

# INVESTIGATION OF POLYMORPHISMS IN SCHIZOPHRENIA RELEVANT GENES

KEREN LEVIEL

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Genetics

Chapel Hill  
2006

Approved by  
Advisor: Patrick F. Sullivan  
Reader: Beverly Koller  
Reader: Ethan Lange  
Reader: Karen Mohlke  
Reader: Fernando Pardo Manuel de Villena

## ABSTRACT

KEREN LEVIEL: Investigation of Polymorphisms in Schizophrenia Relevant Genes  
(Under the direction of Patrick F. Sullivan)

Schizophrenia is a severe, debilitating, and complex disease. It is characterized by delusions, hallucinations, disorganized speech, and grossly disorganized or catatonic behavior. As a complex disease, schizophrenia is thought to be caused by many genes. In addition the etiology is thought to be heterogeneous across the population of persons suffering from schizophrenia. In an attempt to dissect the heterogeneity of the disease, and investigate the role of candidate genes associated with schizophrenia, several studies were completed. The first was a case only association study of neurocognition and *COMT*, a candidate gene for schizophrenia. This is the largest study of its kind to date. The second study addresses the possibility of predisposition due to differential expression of candidate genes. In this study expression of another candidate gene *PRODH* was assessed. Together these studies aimed at investigating the polymorphisms of candidate genes for schizophrenia.

## ACKNOWLEDGEMENTS

Thank you so much to everyone who helped me to get to this point, without whom this would never have been done in time. Thank you to my committee for looking at this on such short notice: Fernando for all of the guidance and support, especially when I was frustrated, Bev for all her love, support, and understanding, Karen for extremely positive and constructive feedback, Ethan for all the statistical help, and Pat for supporting my project financially and going along with all the ups and downs throughout this process.

Thank you to my loving supportive family, especially Mohan who read this over more than anyone, and was supportive through all the ups and downs. I couldn't have done it without you. Thank you to Dana and Seth for just dealing with me, Anat and mom for all of your encouragement, and my Abba who always said I could do anything I put my mind to.

Thank you to all my friends who helped me relax no matter how stressful thing got. Thank you to Crissy and Demetra for many many many hours at Open Eye café, for always putting a positive spin on everything, and for taking many breaks to listen to me. Thank you to John, Corey and Ari for helping with all of my computer problems, and Corey also for critically reading with me. Thank you to Andrea and Tara who were extremely supportive throughout the process, and always had a shoulder I could cry on. Also Tara was always willing to drop what she was doing to help me out. Thank you to Folami for guidance with the process as a whole. And I couldn't forget Megan, who not only helped me at every turn of this project, but also stuck around to help after leaving UNC, well above and beyond what was expected.

## Table of Contents

Chapter 1 .....	1
Introduction.....	1
Chapter 2 .....	6
Genotyping Accuracy for Whole-Genome Amplification of DNA from Buccal Epithelial Cells .....	6
Introduction.....	6
Methods.....	7
Results.....	8
Discussion .....	9
Chapter 3 .....	12
Negative Association of COMT with Neurocognition in the CATIE Study .....	12
Introduction.....	12
Methods.....	14

Results.....	20
DISCUSSION.....	22
Chapter 4.....	30
Differential Allelic Expression of Candidate Genes For Schizophrenia .....	30
Introduction.....	30
Materials and Methods.....	34
Mouse Breeding.....	34
Mouse Dissection and RNA extraction.....	35
DNA Extraction .....	36
Assay Design .....	36
Standardization .....	38
Results.....	39
Determination of the linear range: .....	40
Equal Amplification:.....	40
DNA contamination: .....	41
Differential Allelic Expression: .....	41

Conclusion: .....	42
Chapter 5 .....	52
Conclusion .....	52
Works Cited .....	54

# LIST OF TABLES AND FIGURES

<b>Table 2.1:</b> Comparison of genotyping accuracy between lymphocyte and amplified buccal DNA.....	11
<b>Figure 3.1:</b> Depiction of the chromosome 22q11 region containing COMT.....	25
<b>Table 3.1:</b> Descriptive data for CATIE subjects.....	26
<b>Table 3.2:</b> CATIE allele frequency data for 10 SNPs in COMT. ....	27
<b>Table 3.3:</b> Patterns of linkage disequilibrium within COMT by inferred ancestry. ....	28
<b>Table 3.4:</b> Single SNP results for phenotype-genotype associations.....	29
<b>Figure 4.1:</b> Aldoa Linear Range .....	44
<b>Figure 4.2:</b> Prodh Linear Range.....	45
<b>Figure 4.3:</b> Graph of Aldoa Linear Range .....	46
<b>Figure 4.4:</b> Graph of Prodh Linear Range .....	47
<b>Figure 4.5:</b> DNA Contamination: .....	49
<b>Figure 4.6:</b> PRODH Expression.....	50
<b>Table 4.1:</b> RNA samples available.....	51

## LIST OF ABBREVIATIONS

APC	adenomatous polyposis coli tumor suppressor gene
bp	base pair
CATIE	Clinical Antipsychotic Trials of Intervention Effectiveness
<i>COMT</i>	catechol-O-methyltransferase
DAE	differential allelic expression
<i>DAT</i>	Dopamine transporter
DISC 1& 2	Disrupted in schizophrenia 1& 2
DRD3	Dopamine receptor type 3
<i>DTNBP1</i>	Dysbindin
FAP	familial adenomatous polyposis
FET	Fisher's Exact Test
Het	Heterozygosity
HWE	Hardy-Weinberg equilibrium
IACUC	Institutional Animal Care and Use Committee



LD	linkage disequilibrium
ng	nanogram
<i>NRG1</i>	Neuregulin
Ns	Not significant
PCR	polymerase chain reaction
<i>PRODH</i>	Proline dehydrogenase 1
<i>RGS4</i>	regulator of g-protein signaling 4
RUCDR	Rutgers University Cell and DNA Repository
<i>SLC6A4</i>	serotonin transporter
SNPs	single nucleotide polymorphisms
SCID	Structured Clinical Interview for DSM-IV
SSCP	single strand conformation polymorphism
VCFS	Velo-Cardio-Facial Syndrome
WCST	Wisconsin Card Sorting Test
WRAT3	Wide Range Achievement Test 3

# Chapter 1

## ***Introduction***

Schizophrenia is a severe mental illness. The current criteria for the diagnosis of schizophrenia according to the DSM-IV (American Psychiatric Association 1994) is

A. *Characteristic symptoms:* Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated):

1. delusions
2. hallucinations
3. disorganized speech (e.g., frequent derailment or incoherence)
4. grossly disorganized or catatonic behavior
5. negative symptoms, i.e., affective flattening, alogia, or avolition

**Note:** Only one Criterion A symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the person's behavior or thoughts, or two or more voices conversing with each other.

B. *Social/occupational dysfunction:* For a significant portion of the time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, failure to achieve expected level of interpersonal, academic, or occupational achievement).

C. *Duration:* Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet Criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the

signs of the disturbance may be manifested by only negative symptoms or two or more symptoms listed in Criterion A present in an attenuated form (e.g., odd beliefs, unusual perceptual experiences).

- D. *Schizoaffective and Mood Disorder exclusion:* Schizoaffective Disorder and Mood Disorder With Psychotic Features have been ruled out because either (1) no Major Depressive, Manic, or Mixed Episodes have occurred concurrently with the active-phase symptoms; or (2) if mood episodes have occurred during active-phase symptoms, their total duration has been brief relative to the duration of the active and residual periods.
- E. *Substance/general medical condition exclusion:* The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.
- F. *Relationship to a Pervasive Developmental Disorder:* If there is a history of Autistic Disorder or another Pervasive Developmental Disorder, the additional diagnosis of Schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month (or less if successfully treated).

It is widely suspected that the etiology of schizophrenia is heterogeneous and the proportion of genetic and or environmental causes of their illness is dissimilar both within and across populations. This heterogenic property makes it very difficult to study schizophrenia, a characteristic of the disease that will be discussed further in Chapter 3 of this thesis. The lifetime prevalence of schizophrenia in the U.S. and in most countries is about 1% (Jablensky, Sartorius et al. 1992), with a few populations such as the Palau Islands a Micronesian population isolate which have an increased prevalence of 2% (Camp, Neuhausen et al. 2001).

Generally the onset of schizophrenia is in early adulthood, between 15 and 25 years of age. The average age of onset tends to be higher for females than for males. In 2005, it was estimated that the societal cost of schizophrenia in 2002 was \$62.7 billion of which \$32.4 billion was mainly due to unemployment (Wu, Birnbaum et al. 2005). Thus, any information which leads to a reduction in the amount of time a person who suffers from schizophrenia spends unemployed, would be beneficial not only to the patient and his family, but also to the U.S. economy.

Several genetic epidemiological studies have shown that schizophrenia has both a genetic and an environmental component. In a 1998 review of family and twin studies published between 1920 and 1987, Moldin (Moldin and Gottesman 1997) found the average concordance rate for schizophrenia among monozygotic twins to be 46% even when reared by different families, while the average concordance rate for dizygotic twins was only 14%. In addition, risk of schizophrenia increases with increasing biological relatedness to a person suffering from schizophrenia. This means that third-degree relatives have a lower risk than second-degree relatives who have a lower risk than first-degree relatives who have a lower risk than monozygotic twins. When biological children of a parent with schizophrenia are adopted, they develop schizophrenia or other related disorders at the increased rate seen in first-degree relatives rather than that of the general population (Kety 1988; Kety, Wender et al. 1994). A recent meta-analysis by Sullivan et al which analyzed all published twin studies of schizophrenia, reported that schizophrenia, while heterogeneous, develops as a result of both genetics and environment (Sullivan, Kendler et al. 2003). Clearly genetics plays a role in schizophrenia, but it does not act alone, if it did monozygotic twins would be 100% concordant rather than the reported 46-48%. This means that other factors such as

environment, X inactivation, or methylation differences may be involved. Since it is quite clear that factors other than genetics are involved, when considering a genetic study, heritability is important. There have been several different estimates of the heritability of schizophrenia ranging from 80 to 86 percent (Farmer, McGuffin et al. 1987; Onstad, Skre et al. 1991; Cannon, Kaprio et al. 1998). The best available estimate is 81% (Sullivan, Kendler et al. 2003). While the reported heritability for schizophrenia is high, it cannot be completely understood without taking environmental factors into account.

Schizophrenia is a complex trait (Sullivan, Kendler et al. 2003), for which there have been many attempts to elucidate its etiology. Attempts include studies of cytogenetics, epigenetics, linkage, association, putative biomarkers, and multiple epidemiological risk factors (e.g., seasonal birth effects, exposures *in utero*, and obstetric complications).

One of the major genetic findings in schizophrenia is its connection with velo-cardio-facial syndrome (VCFS). VCFS is caused by a deletion of a 1.5-3 megabase section of chromosome 22q11. Persons suffering from VCFS have a 1 in 4 chance of developing schizophrenia, which is quite higher than the general population risk of about 1% (Saha, Chant et al. 2005). In addition, approximately 1% of patients with schizophrenia possess the VCFS deletion (Horowitz, Shifman et al. 2005). The set of genes deleted in this syndrome is of great interest to the study of schizophrenia. In addition, as previously mentioned other studies such as association and linkage studies have also lead to a large number of candidate genes for the study of schizophrenia. After reviewing recent meta analyses (Harrison and Weinberger 2005) and other research available, a list of candidate genes (Sullivan 2005) for schizophrenia was generated for the purpose of the work which is to follow. The following genes were included in that list: *AKT1*, *catechol-O-methyltransferase (COMT)*, *disrupted in*

*schizophrenia 1 & 2 (DISC 1 & 2), dopamine transporter (SLC6A3, DAT), dopamine receptor type 3 (DRD3), dysbindin (DTNBP1), neuregulin (NRG1), proline dehydrogenase 1 (PRODH), regulator of G-protein signaling 4 (RGS4), and serotonin transporter (SLC6A4).* Two of these genes (*COMT* and *PRODH*) were investigated in the studies reported in this thesis.

This body of work includes three different investigations aimed to further understand the potential importance of polymorphisms in the 22q11 genes *COMT* and *PRODH*. Both *PRODH* and *COMT* have been linked to schizophrenia. *COMT* has been shown to be strongly associated with schizophrenia, while to date only a weak association to *PRODH* has been reported (Harrison and Weinberger 2005). The first investigation (Chapter 2), is an analysis of a whole genome amplification method. The goal of this study is to determine if the method is accurate. The amplification method could potentially be used to amplify human DNA samples for which very small quantities are available. Many potential control groups for the association study that follows have very small amounts of DNA available. Hence, an accurate method for whole genome amplification would allow for flexibility in finding a control group. The next study (Chapter 3) is aimed at investigating polymorphisms of *COMT* in relation to neurocognition in persons with schizophrenia. Studying neurocognition as an endophenotype for schizophrenia will allow us to dissect the heterogeneity of the disease. The final investigation (Chapter 4) focuses on the expression of the candidate genes. It examines the hypothesis that differences in expression of certain genes may lead to predisposition to disease. So, when considering the variations in the set of candidate genes, expression levels are also relevant. The aim of this chapter is to establish and test the tools necessary to study differential allelic expression (DAE) of the candidate

genes in mice, and attempt to determine if *PRODH* exhibits DAE. Together these three studies set the ground work for further investigation of the polymorphisms of the candidate genes for schizophrenia, and offer some insight into the polymorphisms of *COMT*.

## **Chapter 2**

### ***Genotyping Accuracy for Whole-Genome Amplification of DNA from Buccal Epithelial Cells***

#### ***Introduction***

Many population and twin registries have collected DNA via buccal epithelial cell brush or mouthwash protocols. The choice of this method of DNA collection was due to its low subject burden and relatively small expense (10-20% of the cost of collection of a peripheral venous sample). However, the disadvantage of buccal DNA collection is the lesser quantity and poorer quality of DNA in comparison to DNA derived from lymphocytes. Many investigators now wish to genotype existing samples for a large number of markers and the DNA mass required for these genotyping reactions often exceeds the DNA yield of most buccal DNA extraction protocols. The control samples for our association study come from the Add Health study, for which DNA was extracted from buccal cells, and hence not much DNA is available.

Recently, several whole genome amplification protocols have been described using rolling circle amplification with  $\Phi$ 29 DNA polymerase (Dean, Hosono et al. 2002; Nelson, Cai et al. 2002). These approaches appear to be accurate and robust for DNA from lymphocytes (Hosono, Faruqi et al. 2003; Tranah, Lescault et al. 2003). There are no published data however on how these methods perform with buccal-derived DNA samples. The goal of this



study was to compare the accuracy of whole genome amplification of DNA obtained from buccal samples in relation to lymphocyte DNA from the same individuals.

## **Methods**

Paired DNA samples from 30 individuals (DNA from blood versus amplified buccal DNA) were compared for five microsatellites, two variable number of tandem repeats (VNTRs), and 12 SNPs. All SNPs and one VNTR were genotyped twice to assess replication error. Thirty volunteers ascertained by convenience sampling donated both blood and buccal samples. Genomic DNA was extracted from approximately eight mL of whole blood using a Puregene DNA Purification Kit for whole blood samples (Gentra Systems, Minneapolis, MN). DNA from fresh buccal cells was extracted from three sterile cytology brushes using the Puregene Genomic DNA Purification Kit for buccal samples (Gentra Systems, Minneapolis, MN). Extracted buccal cell DNA (800-1000 ng total yield) was amplified using the GenomiPhi DNA Amplification Kit (Amersham Biosciences) via the manufacturer's protocol and using ~20 ng of DNA to seed the reaction.

Blood and amplified buccal samples from each of the 30 volunteers were genotyped for five microsatellites, two VNTRs, and 12 SNPs. These markers were scattered across the genome, and most are pertinent to neuropsychiatric disorders. The five microsatellites were D10S526, D5S592, FES/FPS, vWA31, and D22S417. The two VNTRs were located in the dopamine transporter (*SLC6A3*, 3' DATVNTR) (Kang, Palmatier et al. 1999) and serotonin transporter (*SLC6A4*, 5-HTTLPR) (Lesch, Bengel et al. 1996). The SNPs were from dbSNP (rs1042713, rs6277, rs6265, rs4680, rs2619539, rs1801282, and rs1801133), ABI "Assay-on-Demand" library (C\_7586657, C\_8878813, C\_304219, and C\_2270166), and deCODE (NRG225133) (Stefansson, Sigurdsson et al. 2002).

All PCR reactions were completed on an MJ Research PTC-200 DNA Engine (Global Medical Instrumentation, Inc, Ramsey, MN). Microsatellite and VNTR markers were run on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). ABI GeneMapper (v3.5) was used for semi-automated genotype calls with verification by an experienced operator and pre-established quality control procedures. SNPs were genotyped using TaqMan pre-developed assay reagents for allelic discrimination and run on an ABI Prism 7900HT Sequence Detection System. Genotypes that could not be easily determined by an experienced operator were termed “no-calls”.

This study was reviewed and approved by the UNC Medical IRB. All subjects provided written informed consent. All samples were anonymized following collection.

## **Results**

We considered three metrics by which to gauge genotyping accuracy (Table 2.1): no-call proportions, reproducibility (test-retest), and concordance (blood versus amplified buccal DNA) (Oliphant, Barker et al. 2002). No-calls were considered missing data for all reproducibility and concordance calculations.

No-Call Proportions. For SNPs, the overall no-call proportions were 0.28% (2/720) for DNA from lymphocytes and 0.69% (n=5/720) for DNA from buccal cells (Fisher’s Exact Test  $p=0.45$ ). For the remaining genotypes, the overall no-call proportions were 1.25% (3/240) for DNA from lymphocytes and 2.50% (n=6/240) for DNA from buccal cells (Fisher’s Exact Test  $p=0.50$ ). No individual sample was consistently not called.

Reproducibility was determined by repeating genotyping for 12 SNPS and one VNTR (DATVNTR). Genotype calling was blind to previous results. SNP reproducibility was 100%

(358/358) for DNA from blood and 99.15% (352/355) for amplified buccal DNA (Fisher's Exact Test  $p=0.12$ ). For the one VNTR, reproducibility was 100% for DNA from blood (29/29) and amplified buccal DNA (27/27) (Fisher's Exact Test  $p=1$ ).

Concordance was assessed by pairwise comparison of blood versus amplified buccal DNA from the same subjects. For the 12 SNPs, the concordance was 99.58% (710/713). For the remaining markers, concordance was 100% (232/232).

## ***Discussion***

Our results suggest that whole genome amplification of buccal DNA samples via a  $\Phi$ 29 DNA polymerase/rolling circle method produces genotypes that are of comparable quality to those from genomic DNA from lymphocytes. Specifically, for 12 SNPs, five microsatellites, and two VNTRs, the no-call proportions and reproducibility were not significantly different for lymphocyte and amplified buccal DNA. In addition, the pairwise concordance for blood and amplified buccal samples was very high.

We thought it reasonable to consider no-call genotypes as missing for two reasons. First, inspection of the dataset strongly suggested that these were missing at random with respect to the individual DNA samples. Second, these no-calls can generally be resolved with additional genotyping.

These results do not remove the necessity to evaluate the adequacy of buccal whole genome amplification for every marker assessed. For example, markers requiring long PCR may be more likely to fail on buccal DNA (Roberts, Joyce et al. 2000). Moreover, even if whole genome amplification works well on average across the genome, there are likely to be discrete regions or markers for which it works far less well.

However, if investigators use appropriate experimental and design precautions, our results suggest that  $\Phi$ 29 DNA polymerase whole genome amplification can be used to increase the numbers of markers that can be genotyped on DNA from buccal samples collected from twin and other large population registries.

This is a slightly modified version of a published manuscript (Leviel, Olarte et al. 2004). All statistical analysis was performed by Patrick F. Sullivan. I gathered subjects, collected samples, performed the amplifications, and completed the genotyping.

**Table 2.1:** Comparison of genotyping accuracy between lymphocyte and amplified buccal DNA.

Metric	SNPs		Microsatellites/VNTRs	
	Lymphocyte	Buccal	Lymphocyte	Buccal
Percent no calls	0.28%	0.69%	1.25%	2.50%
Reproducibility <sup>1</sup>	100%	99.15%	100%	100%
Concordance	99.58%		100%	

<sup>1</sup>Microsatellite/VNTR reproducibility data includes DATVNTR only.

## **Chapter 3**

### ***Negative Association of COMT with Neurocognition in the CATIE Study***

#### ***Introduction***

A promising but underutilized approach to studying schizophrenia has been the study of endophenotypes. Endophenotypes are hypothesized to index disease liability more directly, as they are often quantitative traits and likely carry increased statistical power. This approach can also be viewed as a way to dissect the clinical heterogeneity of schizophrenia.

Neurocognitive functions are cognitive functions such as problem solving and language production which are closely related to regions of the nervous system.

Neurocognition is a plausible endophenotype for schizophrenia (Gottesman and Gould 2003). Indeed, neurocognitive dysfunction may be a defining feature of schizophrenia (Elvevag and Goldberg 2000) and cognitive impairment may be the best predictor of poor outcome (Green 1996). Following the Gottesman and Gould definition of an endophenotype (Gottesman and Gould 2003): a) Neurocognition is associated with schizophrenia in the general population with multiple domains  $>1$  SD lower in individuals with schizophrenia than healthy controls (Saykin, Gur et al. 1991; Heinrichs and Zakzanis 1998). Patients with more severe cognitive deficits tend to have more negative symptoms (Manschreck, Maher et al. 1985; Addington, Addington et al. 1991; Strauss 1993), disorganization (Spitzer 1993), and adaptive dysfunction (Green 1996); b) Multiple domains of neurocognition are heritable – e.g., executive function (Swan and Carmelli 2002; Anokhin, Heath et al. 2003), working

memory (Cannon, Huttunen et al. 2000; Tuulio-Henriksson, Haukka et al. 2002), verbal learning and memory (Bouchard 1998; Tuulio-Henriksson, Haukka et al. 2002; Alfimova and Uvarova 2003), vigilance (Myles-Worsley and Coon 1997; Cannon, Huttunen et al. 2000), and processing speed (Swan, Carmelli et al. 1990; Swan and Carmelli 2002; Francks, Fisher et al. 2003; Luciano, Wright et al. 2004); c) Neurocognitive impairment is primarily independent of clinical state – present at first episode of psychosis, and often as severely impaired at first episode as in chronic schizophrenia (Hoff, Riordan et al. 1992; Saykin, Shtasel et al. 1994; Bilder, Goldman et al. 2000); d) Many domains of neurocognition co-segregate with schizophrenia in families (Franke, Maier et al. 1992; Chen, Liu et al. 1998; Cannon, Huttunen et al. 2000; Michie, Kent et al. 2000; Egan, Goldberg et al. 2001); and e) Neurocognitive impairment is more common in unaffected family members than the general population (Cornblatt, Risch et al. 1988; Roxborough, Muir et al. 1993; Keefe, Silverman et al. 1994; Park, Holzman et al. 1995; Kremen, Faraone et al. 1998; Faraone, Seidman et al. 1999; Laurent, Moreaud et al. 1999).

*COMT* (22q11.2) encodes a methyltransferase involved in dopamine metabolism. It affects prefrontal cortical function via modulation of neuronal dopamine activity in the prefrontal cortex and thereby affects neuronal response during working memory. A functional polymorphism (valine → methionine (Val/Met)) has been identified and characterized (Chen, Lipska et al. 2004). Egan and colleagues (Egan, Goldberg et al. 2001) reported an association between the more active Val allele and lower scores on the Wisconsin Card Sorting Test (WCST), which supports the idea that higher dopamine levels lead to better performance in executive functions. Among schizophrenia patients the association with the polymorphism has been very inconsistent (Harrison and Weinberger 2005) as is its

association with neurocognitive performance. Many of the recent publications which address the role of *COMT* in relation to neurocognition tend to have small sample sizes and usually focus only on the functional Val/Met polymorphism (Harrison and Weinberger 2005). This has been studied in a variety of populations including persons suffering from schizophrenia, their normal siblings, and persons suffering from ADHD (Bellgrove, Domschke et al. 2005; Harrison and Weinberger 2005; Turic, Williams et al. 2005). In one of these studies an association was found between the normal siblings of persons suffering from schizophrenia, but not in the patients (Rosa, Peralta et al. 2004). In addition Shifman et al., found a weak association with this allele, but a highly significant association with a *COMT* haplotype (Shifman, Bronstein et al. 2002). These types of results exemplify the need for a larger sample size in association studies, in addition to the need for more than one SNP to be genotyped.

The goal of this investigation was to examine the associations of neurocognitive scores with 10 single nucleotide polymorphisms (SNPs) in or near the *COMT* gene (including the widely studied val158met SNP rs4680). The sample was drawn from the CATIE study and, to our knowledge, is the largest sample in which the molecular genetics of neurocognition has been investigated in schizophrenia.

## **Methods**

Subjects. All subject were participants in the Clinical Antipsychotic Trials of Intervention Effectiveness project (CATIE, NIMH contract NO1 MH90001) which was conducted between January 2001 and December 2004. CATIE was a multi-phase randomized controlled trial of antipsychotic medications involving 1,460 persons with schizophrenia



followed for up to 18 months (Stroup, McEvoy et al. 2003; Lieberman, Stroup et al. 2005). The philosophy of the trial was to assess controlled treatment with antipsychotic drugs in a broad range of patients with schizophrenia under “real world” conditions. To maximize the representativeness of the sample, subjects were ascertained from a broad array of clinical settings scattered across the US (16 university clinics, 10 state mental health agencies, 7 Veterans Affairs medical centers, 6 private nonprofit agencies, 4 private-practice sites, and 14 mixed-system sites). 1,894 subjects were evaluated and 1,460 (77.0%) entered into CATIE. No subject was known to be related to any other subject. All subjects provided written informed consent (including an additional consent for genetic studies), and the full study protocol was reviewed by internal review boards at UNC and at participating study sites.

Preliminary diagnoses were established by referring psychiatrists. Final study diagnoses were independently re-evaluated by CATIE personnel using the Structured Clinical Interview for DSM-IV (SCID) (First, Spitzer et al. 1994) using all available information (including psychiatric and general medical records) along with one or more subject interviews. Interviewers were experienced Master’s-level clinicians who were specifically trained to criterion via a standard protocol (First, Spitzer et al. 1994). Any diagnostic uncertainties were resolved via discussion with one of the CATIE senior clinicians. The DSM-IV (Association 1994) criteria for schizophrenia operationalized in the SCID include the explicit requirement for characteristic symptoms, significant dysfunction, and continuous signs of illness for  $\geq 6$  months. The SCID explicitly excludes clinically important phenocopies—e.g., mood disorders with psychotic features or symptoms resulting from the direct physiological effects of a substance (e.g., drugs of abuse or medications) or a

general medical condition. The definition is inherently rigorous and requires substantial evidence that consistently supports the diagnosis.

*Inclusion criteria* were: definite diagnosis of schizophrenia (American Psychiatric Association 1994; First, Spitzer et al. 1994), age 18-65 years, clinical decision that oral medication is appropriate, adequate decisional capacity, and provision of written informed consent. *Exclusion criteria* – see Table 2 in Stroup et al. for a complete list (Stroup, McEvoy et al. 2003). Briefly, patients were excluded if they had received a diagnosis of schizoaffective disorder, mental retardation, or other cognitive disorders; had a history of serious adverse reactions to the proposed treatments; had had only one schizophrenic episode; had a history of treatment resistance, defined by the persistence of severe symptoms despite adequate trials of one of the proposed treatments or prior treatment with clozapine; were pregnant or breastfeeding; or had a serious and unstable medical condition. Individuals with psychoactive drug use disorders were included but only when there was positive evidence that schizophrenia was an independent diagnosis. The study was approved by the institutional review board at each site, and written informed consent was obtained from the patients or their legal guardians.

DNA Sampling & Cell Line Establishment. Peripheral venous blood samples were obtained and sent to the Rutgers University Cell and DNA Repository (RUCDR) where cell lines were established via Epstein-Bar Virus transformation. Numerous quality control procedures are routine and the success rate for immortalization exceeds 99% ([http://www.rucdr.org/quality\\_control.html](http://www.rucdr.org/quality_control.html)). Sample DNA concentrations were quantified and normalized via the use of Picogreen dsDNA Quantitation Kits (Molecular Probes, Eugene, OR).

SNP Selection, Genotyping, & Quality Control. Ten SNPs were genotyped including rs6518592, rs2097603, rs2020917, rs737865, rs740603, rs740601, rs4680, rs4646316, rs165774, and rs165599. rs6518592, rs2020917, and rs740603 were chosen for genotyping because they represent a HapMap haplotype in the Caucasian population, as were rs740601, rs4680, rs4646316, and rs165774. Those SNPs were chosen in 2004 when this project was designed, and were chosen because they are tagging SNPs. Three SNPs, including rs737865, rs4680 (part of another haplotype, and coding SNP), and rs165599 were chosen because of past research by Shifman et al, which found a significant association of this haplotype with schizophrenia. The last SNP rs740603, was chosen because of its location in the promoter region of the gene. In all these represent 2 HapMap haplotypes (release #11, 9/04), a haplotype previously reported by Shifman et al., and an additional promoter SNP.

Representing the largest number of SNPs genotyped in any *COMT* study.

We used TaqMan 5' exonuclease assays (Livak 1999; Chen and Sullivan 2003) to genotype these 10 SNPs with probes and primers designed and supplied by Applied Biosystems (Foster City, CA). SNP markers were genotyped in a 384-well format. Each well contained 5.0 ng of genomic DNA, two allele-specific fluorescent probes, and ABI TaqMan Universal Master Mix without UNG. The total reaction volume was 5  $\mu$ L with PCR primer concentrations of 900 nM and TaqMan MGB-probe concentrations of 200 nM. PCR thermocycling conditions were: 95° C for 10 minutes, followed by 50 cycles of 95° C for 15 seconds and 60° C for 1 minute, with a final incubation at 25° C (PE 9700, ABI).

Fluorescence intensities were read by an ABI 7900HT. Automated allele calls were made with SDS Data Collection software with all allele calls reviewed by an experienced operator according to protocol. To pass quality control, we required that each plate have: no-call rates

≤5.0%, a random pattern of no-calls, and all genotypes in Hardy-Weinberg equilibrium (HWE) when stratified by inferred ancestry. The overall no-call percentage was 1.3% (82/6,336). Agreement between 48 duplicate samples per SNP was 100%. Genotyping was conducted blind to all clinical data.

Neurocognitive Phenotypes. Analyses in this report are from the CATIE baseline dataset. The construction, implementation, and analyses of the neurocognitive data from CATIE are described at length elsewhere (Keefe, Mohs et al. 2003; Keefe, Bilder et al. Submitted). Briefly, this battery assessed aspects of neurocognitive function that are robustly associated with schizophrenia, that could plausibly improve with treatment, and which were practical to do in a large patient sample treated in “real world” clinical care settings distributed across the US. Despite minimal screening procedures, 91.2% of patients provided meaningful neurocognitive data at baseline entry into CATIE. There were 24 individual scores from 11 neurocognitive tests that were grouped into five domain scores (processing speed, reasoning, verbal memory, working memory, and vigilance). As detailed elsewhere (Keefe, Bilder et al. Submitted) and following the MATRICS factor model (Nuechterlein, Barch et al. 2004), a composite neurocognitive score (standardized average of the five domain scores) was determined to be a reasonable empirical summary of these data and was chosen *a priori* as the dependent analytic variable.

We used multiple regression with stepwise selection to evaluate the multivariate correlates of the composite neurocognitive measure. Of a set of 16 potential covariates (none of which included genetic data), better composite neurocognitive scores were significantly associated with seven empirical covariates: higher scores on the reading subtest of the WRAT-3 ( $p<0.00005$ ), lower PANSS negative symptom scale scores ( $p<0.00005$ ), fewer

years since the age at which an antipsychotic was first prescribed ( $p < 0.00005$ ), younger age at CATIE baseline ( $p = 0.0002$ ), greater years of education ( $p = 0.0002$ ), the absence of a recent clinical exacerbation ( $p = 0.002$ ), and the absence of evidence of clinically significant drug use ( $p = 0.01$ ). These variables were used as covariates in adjusted analyses.

Population stratification is a potential limitation of association studies (Risch 2000; Sullivan, Eaves et al. 2001) and is a salient concern in CATIE given its explicit intent to ascertain a diverse, clinically representative sample (Stroup, McEvoy et al. 2003; Lieberman, Stroup et al. 2005). Stratification effects are more likely with fewer strata and larger samples (Devlin and Roeder 1999; Wacholder, Rothman et al. 2000; Freedman, Reich et al. 2004). Some have argued that the threat of population stratification is exaggerated (Risch and Teng 1998; Risch 2000) and that controlling for reported ancestry will reduce the risk (Devlin and Roeder 1999; Wacholder, Rothman et al. 2000). There are two preconditions for population stratification to yield false positive results (Gorroochurn, Hodge et al. 2004; Heiman, Hodge et al. 2004). Because population stratification as a confounder exists only if the phenotypic distribution *and* the marker allele frequency differ across ancestral strata, we will be able to determine the impact of stratification by measuring the marker and disease frequency and predict the extent of stratification. If there is substantial evidence that population substructure exists in the overall CATIE sample for the markers studied here, we default to stratified analysis. CATIE subjects were allowed to select multiple racial categories (White, Black/African-American, American Indian or Alaska native, Asian, Native Hawaiian or Pacific Islander, or Other) and were also asked if they were Hispanic or Latino. Inferred ancestry was defined as “Africa only” if a subject endorsed Black/African-American only, “Europe only” if a subject endorsed White only, and “Other” if any other racial category

was endorsed. Our rationale for use of inferred ancestry as an index of population stratification is described at length elsewhere (Shields, Fortun et al. 2005).

Statistical Analysis. The logic underlying our analyses is explicated in the Results. Data management and statistical analyses were performed with SAS (version 9.1) (SAS Institute Inc. 2004). Significance testing and adjustment for multiple comparisons were via a permutation approach (Efron and Tibshirani 1993; Lazzeroni and Lange 1998; Good 2004). HaploView (version 3.2) (Barrett, Fry et al. 2004) was used to test Hardy-Weinberg Equilibrium (HWE) and  $|D'|$  and  $r^2$  linkage disequilibrium estimates (Devlin and Risch 1995). Haplotype analyses were conducted with the R (version 2.1.1) (R Development Core Team 2005) function haplo.score (contained in version 1.2.0 of the haplo.stats library) (Schaid, Rowland et al. 2002).

Project Context. All COMT SNPs genotyped are reported. Additional genes are under investigation (see <http://www.med.unc.edu/~pfsulliv/downloads.htm> for an updated list). We have attempted to follow published guidelines for association studies (Sullivan, Eaves et al. 2001; Little, Bradley et al. 2002; Colhoun, McKeigue et al. 2003).

## **Results**

In this study, DNA samples from 641 CATIE participants were used. Descriptive information for these subjects is detailed in Table 3.1. In general, this population was intended to be diverse, with a chronic course of disease. The people included in the CATIE study have suffered from schizophrenia for a long time, with a mean time since first antipsychotic prescribed of 14.16 years. Of the 641 subjects, 450 (73.3%) are males and only 168 are females.

The region of chromosome 22q11.21 including *COMT* is depicted in Figure 3.1. Each of the 10 SNPs genotyped are noted, including the haplotype from Shifman et al. (Shifman, Bronstein et al. 2002) the Val158Met (rs4680) functional polymorphism, and the promoter SNP.

There are two reasons which may lead one to stratify by race to avoid false positives. The first is if the phenotypic distribution differs across strata. Initially there did seem to be a difference in the phenotypic distribution of perseverative errors or neurocognitive composite score by inferred ancestry. However, once the composite neurocognitive score or perseverate error score was adjusted for empirical covariates, there was no association with inferred ancestry. The composite neurocognitive score was adjusted for seven empirical covariates which included the reading subtest of the WRAT-3, PANSS negative symptom scale scores, years since an antipsychotic was first prescribed, age, education, any recent exacerbation, and no clinically significant drug use. The perseverative error score was adjusted for four empirical covariates including the reading subtest of the WRAT-3, PANSS negative symptom scale scores, years since an antipsychotic was first prescribed, and Calgary depression score. Since the phenotypic difference seen between the inferred ancestry groups was eliminated after adjusting for the various covariates, the first reason to stratify by inferred ancestry is eliminated.

The second reason for possibly stratifying data is if there is an allele frequency difference across the strata. HWE testing was performed for each SNP for all samples and by strata (Table 3.2). Three SNPs were out of HWE when looking at the entire sample, however when looking at the allele frequencies for each of the three SNPs (rs740601, rs4680, rs165599), and testing for HWE for each SNP by inferred ancestry, all were in HWE. Since

the phenotypic distribution across strata was not significantly different, association was tested without stratification.

Linkage disequilibrium testing was performed by strata,  $D'$  and  $r^2$  values were calculated for each pair of SNPs. Table 3.3 gives both values for each pair of SNPs. Using the  $r^2$  values only shows that relatively few SNPs are in LD. The SNPs chosen do not represent very good coverage of *COMT*, however it is the best to date.

Association of each SNP with each of the two phenotypes was assessed, the results are listed in Table 3.4. For the WCST or perseverative errors, two SNPs (rs651892 and rs4680) were significantly associated at the 0.05 level, however once adjusting for multiple comparisons the association was not statistically significant. Interestingly, the association with the functional SNP (rs4680), was opposite of what was previously reported by Egan et al (Egan, Goldberg et al. 2001). In the previous study by Egan et al, the Met allele was found to enhance prefrontal cognition, while here we find it to impair.

Haplotype analysis was performed for each of the three haplotypes, two HapMap haplotypes and the haplotype identified by Shifman et al. (results not shown). None of the SNPs or haplotypes were found to be significantly associated with either composite neurocognitive score or perseverative error score.

## ***DISCUSSION***

The initial hypothesis of this study, which was designed in 2004, was that rs4680 would be significantly associated with score on the WCST (perseverative error). In addition, it was thought that the Met allele would enhance performance and the Val allele would



impair performance on this test, since this is what was previously reported (Egan, Goldberg et al. 2001). In this study however, no significant association was found, and better performance on the WCST tended to correlate with Val allele. Then two HapMap haplotypes, and a haplotype from Shifman et al.(Shifman, Bronstein et al. 2002) were added in addition to a promoter SNP. None were found to significantly associate with WCST score, even when stratified by sex. The composite neurocognitive score was then added to the analysis, and again no association was observed.

These results represent the largest study to date for association of *COMT* genotypes with perseverative error score or composite neurocognitive score. While no association was observed, this does not mean that *COMT* is not involved in predisposition to schizophrenia. It simply means that the markers which were genotyped, and those in LD with those markers may not be involved in neurocognition. While this is the largest number of SNPs to be genotyped in *COMT*, it does not represent full gene coverage, which would require at least twenty more SNPs, so a SNP which is associated with the neurocognitive phenotype may have been missed. This study was designed before the most recent HapMap information was available. Now that more information is available, future studies could be designed with better coverage of *COMT*.

Since this study was designed, there have been other negative association studies of *COMT* with the WCST (Mills, Langley et al. 2004; Taerk, Grizenko et al. 2004). It is possible that the previous finding of association was a false positive, or that performance on the WCST varies and the subjects in the CATIE trial were only assessed once. It is also important to note that in this study an endophenotype is being measured rather than the presence or absence of schizophrenia, so *COMT* may still be involved in some other aspect

of the disease not related to neurocognition. Further studies are necessary including case-control association studies, and the CATIE sample is an excellent population to perform future analyses.

All of the statistical analysis mentioned in this chapter was performed by Patrick F. Sullivan. I performed all molecular analysis, organized samples, performed all genotyping and genotyping quality control.



**Table 3.1:** Descriptive data for CATIE subjects <sup>1</sup>.

Continuous variables	Mean	SD	Median	IQR	Range
Age in years	40.7	11.1	42	33, 49	18, 67
Years since first antipsychotic prescribed	14.1	10.6	13	4, 22	0, 56
Years of formal education	12.2	2.2	12	11, 13	3, 21
Clinician global severity score	3.9	1.0	4	3, 4	1, 7
PANSS total score	73.1	17.5	73	61, 85	33, 132
PANSS positive symptom score	17.8	5.6	18	14, 22	7, 36
PANSS negative symptom score	19.7	6.5	20	15, 24	7, 41
WCST perseverative errors <sup>2</sup>	0.0	1.0	0.31	-0.42	0.66
Neurocognitive composite score <sup>2</sup>	0.0	0.9	0.02	-0.63	0.63
Nominal variables	Number	Percent			
Male gender	450	73.3%			
Inferred continental ancestry					
Africa only	177	28.8%			
Europe only	365	59.4%			
Other	72	11.7%			
Married or with significant other	64	10.4%			
Current employment, full or part-time	39	6.4%			
Hospitalized/received crisis treatment in prior 3 months	158	25.7%			
First language is English	574	93.5%			
Lifetime drug/alcohol abuse or dependence (SCID)	281	45.8%			
Current drug/alcohol abuse or dependence (SCID, clinician, or hair toxicology)	183	29.8%			

<sup>1</sup> Abbreviations. SD=standard deviation. IQR=interquartile range, p25-p75. WCST=Wisconsin Card Sorting Test.

<sup>2</sup> Directionality: greater scores mean better performance.

**Table 3.2:** CATIE allele frequency data for 10 SNPs in COMT.

Chr. 22 Position	Locus (rationale)	Inferred Ancestry	N	HWE		Allele 1		Allele 2	
				$\chi^2$	P value	Base	Freq.	Base	Freq.
18,301,700	Rs6518592 (htSNP, block 1)	Entire sample	695	1.02	Ns	A	0.37	C	0.63
		European only	395	0.00	Ns		0.42		0.58
		African only	208	2.66	Ns		0.27		0.73
		Other	92	0.29	Ns		0.36		0.64
18,302,646	Rs2097603 (P2 promoter)	Entire sample	696	0.07	Ns	A	0.60	G	0.40
		European only	395	0.16	Ns		0.58		0.42
		African only	209	0.27	Ns		0.64		0.36
		Other	92	0.74	Ns		0.63		0.38
18,303,438	Rs2020917 (htSNP, block 1)	Entire sample	696	0.43	Ns	C	0.76	T	0.24
		European only	395	0.11	Ns		0.69		0.31
		African only	209	0.93	Ns		0.89		0.11
		Other	92	0.01	Ns		0.77		0.23
18,304,675	Rs737865 (Shifman/Bray)	Entire sample	693	0.77	Ns	A	0.75	G	0.25
		European only	393	0.10	Ns		0.70		0.30
		African only	208	0.88	Ns		0.86		0.14
		Other	92	0.08	Ns		0.76		0.24
18,319,731	Rs740603 (htSNP, block 1)	Entire sample	695	0.61	Ns	A	0.51	G	0.49
		European only	394	0.45	Ns		0.48		0.52
		African only	209	0.00	Ns		0.54		0.46
		Other	92	0.10	Ns		0.59		0.41
18,325,317	Rs740601 (htSNP, block 2)	Entire sample	696	5.17	0.02	A	0.62	C	0.38
		European only	395	3.27	0.07		0.62		0.38
		African only	209	2.57	0.11		0.57		0.43
		Other	92	0.48	0.49		0.72		0.28
18,325,825	Rs4680 (htSNP, block 2, Val158met)	Entire sample	692	4.71	0.03	A	0.46	G	0.54
		European only	392	2.66	0.10		0.53		0.47
		African only	208	1.18	0.28		0.31		0.69
		Other	92	2.62	0.11		0.44		0.56
18,326,686	Rs4646316 (htSNP, block 2)	Entire sample	695	0.11	Ns	C	0.80	T	0.20
		European only	394	0.21	Ns		0.79		0.21
		African only	209	0.80	Ns		0.80		0.20
		Other	92	0.46	Ns		0.82		0.18
18,327,115	Rs165774 (htSNP, block 2)	Entire sample	694	3.85	Ns	A	0.29	G	0.71
		European only	393	2.04	Ns		0.35		0.65
		African only	209	0.02	Ns		0.22		0.78
		Other	92	0.16	Ns		0.22		0.78
18,331,335	Rs165599 (Shifman/Bray)	Entire sample	696	10.58	0.001	C	0.44	T	0.56
		European only	395	0.02	0.88		0.29		0.71
		African only	209	1.51	0.22		0.68		0.32
		Other	92	2.06	0.15		0.47		0.53

Het=heterozygosity, HWE=Hardy-Weinberg equilibrium, Ns= Not significant at 0.05 level

**Table 3.3:** Patterns of linkage disequilibrium within COMT by inferred ancestry.

<i>African only</i>	rs6518592	rs2097603	rs2020917	rs737865	rs740603	rs740601	rs4680	rs4646316	rs165774	rs165599
Rs6518592		0.62	0.04	0.06	0.23	0.01	0.07	0.01	0.01	0.00
Rs2097603	0.97		0.07	0.06	0.17	0.00	0.02	0.02	0.00	0.00
Rs2020917	1.00	1.00		0.68	0.00	0.05	0.04	0.01	0.02	0.01
Rs737865	1.00	0.82	1.00		0.01	0.03	0.05	0.01	0.02	0.00
Rs740603	0.85	0.59	0.17	0.17		0.00	0.02	0.05	0.00	0.00
Rs740601	0.15	0.10	0.54	0.39	0.04		0.25	0.07	0.14	0.03
Rs4680	0.30	0.17	0.83	0.81	0.24	0.86		0.10	0.60	0.19
Rs4646316	0.30	0.39	0.11	0.12	0.44	0.45	0.93		0.07	0.00
Rs165774	0.10	0.03	0.77	0.71	0.14	0.82	1.00	1.00		0.29
Rs165599	0.06	0.01	0.14	0.14	0.00	0.31	0.43	0.09	0.70	
<i>European only</i>	rs6518592	rs2097603	rs2020917	rs737865	rs740603	rs740601	rs4680	rs4646316	rs165774	rs165599
Rs6518592		0.98	0.32	0.31	0.54	0.14	0.17	0.03	0.05	0.03
Rs2097603	0.99		0.32	0.31	0.55	0.14	0.17	0.03	0.05	0.04
Rs2020917	1.00	1.00		0.97	0.26	0.40	0.29	0.11	0.14	0.03
Rs737865	1.00	1.00	0.99		0.27	0.39	0.28	0.12	0.14	0.03
Rs740603	0.84	0.84	0.80	0.82		0.28	0.23	0.05	0.09	0.03
Rs740601	0.56	0.55	0.75	0.75	0.70		0.71	0.31	0.34	0.05
Rs4680	0.53	0.52	0.75	0.75	0.54	0.99		0.20	0.46	0.21
Rs4646316	0.39	0.39	0.44	0.44	0.46	0.85	0.80		0.15	0.08
Rs165774	0.25	0.24	0.76	0.75	0.39	0.99	0.98	1.00		0.20
Rs165599	0.34	0.35	0.17	0.17	0.28	0.28	0.66	0.83	0.92	
<i>Other</i>	rs6518592	rs2097603	rs2020917	rs737865	rs740603	rs740601	rs4680	rs4646316	rs165774	rs165599
Rs6518592		0.95	0.17	0.19	0.23	0.06	0.29	0.03	0.00	0.08
Rs2097603	1.00		0.18	0.19	0.21	0.05	0.25	0.03	0.00	0.07
Rs2020917	1.00	1.00		0.91	0.18	0.41	0.10	0.15	0.01	0.00
Rs737865	1.00	1.00	1.00		0.15	0.39	0.12	0.19	0.01	0.00
Rs740603	0.76	0.72	0.64	0.57		0.21	0.03	0.08	0.02	0.01
Rs740601	0.53	0.45	0.74	0.69	0.60		0.31	0.40	0.11	0.00
Rs4680	0.64	0.58	0.67	0.69	0.24	1.00		0.17	0.35	0.28
Rs4646316	0.46	0.45	0.46	0.53	0.49	0.85	1.00		0.06	0.03
Rs165774	0.10	0.06	0.32	0.34	0.20	1.00	1.00	1.00		0.15
Rs165599	0.39	0.36	0.04	0.08	0.15	0.02	0.63	0.36	0.76	

Values below the diagonal are  $D'$  and  $r^2$  above the diagonal. Areas of increased LD are highlighted. Outlined SNP names form haplotype blocks.

**Table 3.4:** Single SNP results for phenotype-genotype associations.

Phenotype	SNP	F	P-value	Genotypes			Least square means		
Aim 1—WCST perseverative errors	rs6518592	3.50	0.0309	AA	AC	CC	-0.24	0.05	0.05
	rs2097603	2.87	0.0575	AA	AG	GG	0.05	0.05	-0.20
	rs2020917	0.97	0.3794	CC	TC	TT	-0.03	0.08	0.10
	rs737865	1.25	0.2859	AA	AG	GG	-0.03	0.07	0.19
	rs740603	1.90	0.1498	AA	AG	GG	-0.10	0.08	0.02
	rs740601	0.99	0.3724	AA	AC	CC	-0.05	0.07	0.04
	rs4680	4.17	0.0160	AA	AG	GG	-0.18	0.10	0.00
	rs4646316	0.19	0.8278	CC	CT	TT	0.00	0.04	-0.08
	rs165774	0.10	0.9067	AA	AG	GG	0.06	0.02	0.00
	rs165599	0.24	0.7834	CC	CT	TT	0.02	0.04	-0.03
Aim2—Composite score	rs6518592	0.16	0.8486	AA	AC	CC	-0.02	0.00	0.01
	rs2097603	0.17	0.8431	AA	AG	GG	0.02	0.00	-0.01
	rs2020917	0.98	0.3756	CC	TC	TT	-0.02	0.04	0.07
	rs737865	1.27	0.2805	AA	AG	GG	-0.02	0.04	0.07
	rs740603	1.02	0.3603	AA	AG	GG	-0.04	0.03	0.02
	rs740601	0.56	0.5702	AA	AC	CC	-0.02	0.03	-0.01
	rs4680	1.28	0.2792	AA	AG	GG	-0.05	0.03	0.00
	rs4646316	0.22	0.8023	CC	CT	TT	0.00	0.01	0.07
	rs165774	0.96	0.3847	AA	AG	GG	0.06	0.03	-0.02
	rs165599	0.03	0.9737	CC	CT	TT	0.00	0.01	0.01

## Chapter 4

# Differential Allelic Expression of Candidate Genes For Schizophrenia

### *Introduction*

Gene variants which lead to changes in gene expression, otherwise known as differential allelic expression (DAE) (Knight 2004), may ultimately result in predisposition to disease. For instance, a specific allele of the adenomatous polyposis coli tumor suppressor gene (*APC*) decreases expression, and ultimately, foments predisposition to the development of familial adenomatous polyposis (FAP) (Yan, Dobbie et al. 2002). Another example of the role of DAE in human diseases is obesity. A T/G polymorphism in the adiponectin gene was reported to be associated with obesity (Stumvoll, Tschritter et al. 2002). A recent study examined the expression of the two alleles in adipose tissue of heterozygotes (Yang, Tsou et al. 2003), and found that the allele associated with obesity showed higher expression. These two examples suggest how DAE could play a role in predisposition to complex diseases.

Since many genes are thought to be involved in predisposition to schizophrenia, the study of DAE for the relevant candidate genes could prove to be a powerful tool. To date, studies have focused predominantly on identifying polymorphisms which result in change in an amino acid sequence. However, it is important to explore the role of non-coding as well as the coding polymorphisms when studying changes in gene expression. In addition, one must also individually examine *cis*-elements or *trans*-acting factors as one might dilute the



signal from the other (Pastinen, Ge et al. 2005), in trying to identify both simultaneously. In this chapter, we are principally interested in *cis*-elements that alter gene expression.

Several studies have addressed the question of whether there are differences in candidate gene expression between persons with and without schizophrenia (Bray, Buckland et al. 2003; Weickert, Straub et al. 2004). In those studies, SNPs that lead to differential expression were assessed in relation to schizophrenia. The role of one candidate gene (*COMT*) has been examined in the following three reports, however with inconsistent results. The first study by Bray et al. assessed *COMT* mRNA expression in post-mortem brains of individuals with and without schizophrenia (Bray, Buckland et al. 2003). They found a significant decrease in *COMT* gene expression associated with a previously identified risk haplotype (rs737865-rs165599) (Shifman, Bronstein et al. 2002). The second study by Tunbridge et al. looked at *COMT* mRNA levels in post-mortem brains from cases with schizophrenia and matched controls and found no difference in *COMT* mRNA abundance between cases and controls, nor did they find an association between mRNA level and *COMT* genotype (Tunbridge, Burnet et al. 2004). The third study by Chen et al. examining *COMT* mRNA levels in cases with schizophrenia and controls found no difference in *COMT* expression, nor did it find an association of any SNP with differential expression (Chen, Lipska et al. 2004). Chen et al. did, however, find a difference in *COMT* protein levels and activity. These studies highlight the inconsistencies in DAE studies in humans with only one of the three studies finding differential *COMT* expression and SNPs that correlate with the difference in expression, while the two studies did not find a significant change in *COMT* expression.

The inconsistent *COMT* expression findings in humans described above may be caused by a variety of factors, four possibilities follow. First, tissue collection in humans presents challenges, because it is imperative to get all samples at the same stage of the disorder, at the same age, and with similar cause of death since gene expression may vary as a result of any of these factors. Since brain tissue can generally only be collected after death in humans, different specimens would not necessarily be of the same age, cause of death and stage of disease. Second, genetic background is a major problem in identifying variants that alter gene expression. One issue with the genetic environment pertains to the difference between identifying *cis*-regulatory elements versus *trans*-acting factors. A difference in gene expression can be due to a change in a *cis*-element or in a *trans*-acting factor neither of which are the same across all humans, and are therefore difficult to isolate (Pastinen, Ge et al. 2005). The third issue is causality. In humans, where the genetic background is different, it is very difficult to distinguish cause (a SNP indeed causes DAE) from a confounding effect (an unmeasured causative SNP is in linkage disequilibrium (LD) with the measured SNP). The fourth issue pertains to environmental effects that are obviously nearly impossible to control for in humans. Subjects studied generally come from various geographic locations, different familial backgrounds, and different environmental exposures any one of which could possibly influence gene expression.

When it is not simple to study a genetic process in humans, a logical step is to use a model organism like inbred mice as all of the difficulties described above could, in theory, be remedied. In addition, mouse models provide the advantage that one could tightly control for environment, genetic background, and age, and eventually, even determine causality. However, even in mice, discerning the extent of *cis*- vs. *trans*-acting regulatory variation can

present difficulties. As with humans, the issue of *trans*-acting factors or genetic environment could affect expression levels (Cowles, Hirschhorn et al. 2002). One cannot simply quantify total gene expression at a locus in two mouse strains, and assume that any differences are due to a *cis*-acting elements. All *trans*-acting factors that may act on the *cis*-element of interest need to be identical for each allele studied. To overcome this difficulty F<sub>1</sub>s could be used. F<sub>1</sub>s are ideal because they carry each of the two alleles of interest within each cell, and all the *trans*-acting factors and environmental conditions are identical. In theory, for each gene of interest, the level of expression of each parental allele would be quantified for each F<sub>1</sub>. Then, F<sub>1</sub>s which show a difference in expression between the two alleles would be used to map the allele/s which cause the change in expression. The steps necessary to complete this type of study are briefly enumerated below.

1. **Panel of mouse strains:** To achieve high mapping resolution a panel of mouse strains was selected. A subset was used as a proof of concept.
2. **Generate F<sub>1</sub>s:** F<sub>1</sub>s were generated from the subset of the pannel.
3. **RNA extraction:** RNA was extracted from each tissue of interest. These will be used to test for DAE.
4. **Quantification of DNA contamination:** The level of DNA contamination across the RNA samples was quantified. DNA may produce a product of the same size as the product from RNA. Adjusting for DNA contamination will allow us to determine how much product is produced from RNA alone.
5. **Assay design:** An assay for each gene of interest had to be designed, tested, and standardized. For the assay to be used, it had to amplify equally across all parental strains, such that differences in expression are measured rather than the

efficiency of the PCR for an allele. The linear range for each assay had to be determined. The expression level should be assessed while the PCR is in its linear phase, before it has reached a plateau. The reason for this is that the amplification of each of the alleles does not plateau at the same time, which means that while one allele is still being doubled the other will not, and the ratio will change. While both are in the linear range, the ratio will be constant, and representative of the initial number of copies of each allele.

6. Finally once an assay is ready, differential expression can be identified for each gene, and potentially causative SNP mapped.

The goal of this study was to create a set of RNA samples, standardize the steps to identifying differential expression, and to assess the presence or absence of differential expression of candidate genes in mouse brain samples.

## ***Materials and Methods***

### **Mouse Breeding**

Mice from the following inbred strains were used: C57BL/6J (B6), A/J, PERA/EiJ (PERA), PERC/EiJ (PERC), CAST/EiJ (CAST), DBA/2J (DBA), WSB/EiJ (WSB), PWD/PhJ (PWD), FVB/NJ (FVB), JF1/Ms (JF1), and RBA/Dn (RBA). All mice were purchased from The Jackson Laboratory (Bar Harbor, ME) with the exception of JF1/Ms which are maintained by Dr. Terry Magnuson at the University of North Carolina at Chapel Hill (UNC-CH). Mice were housed and treated in compliance with the recommendations of the Institutional Animal Care and Use Committee (IACUC) of UNC-CH. The following F<sub>1</sub>s were generated (dam always listed first and sire second) by Timothy A. Bell in the Pardo Manuel de Villena laboratory at UNC-CH: (B6 x A/J) F<sub>1</sub>, (B6 x PERA) F<sub>1</sub>, (CAST x A/J)

F<sub>1</sub>, (DBA x B6) F<sub>1</sub>, (PERA x B6) F<sub>1</sub>, (PERA x WSB) F<sub>1</sub>, (PWD x B6) F<sub>1</sub>, (WSB x B6) F<sub>1</sub>, (A/J x B6) F<sub>1</sub>, (B6 x PWD) F<sub>1</sub>, (B6 x WSB) F<sub>1</sub>, (FVB x B6) F<sub>1</sub>, (JF1 x B6) F<sub>1</sub>, (PWD x A/J) F<sub>1</sub>, (RBA x A/J) F<sub>1</sub>, (PERC x B6) F<sub>1</sub>, (A/J x CAST) F<sub>1</sub>, (FVB x PWD) F<sub>1</sub>.

## **Mouse Dissection and RNA extraction**

Mice were at least 3 months of age prior to harvesting tissues. We extracted RNA from six F<sub>1</sub>s from every cross (3 males and 3 females). All mice were anesthetized using isoflurane (Halocarbon Laboratories, North Augusta, SC), and killed by cervical dislocation. Tissues dissected were brain, cerebellum, heart, lung, colon, liver, kidney, pancreas, testis/ovaries, and spleen. Tissues except for pancreas were placed in TRIzol reagent (Invitrogen, Carlsbad, CA) for same day extraction or, for future extraction, in RNAlater (Qiagen, Valencia, CA) and stored at -80°C. One 3 mm TissueLyser (Qiagen, Valencia, CA) bead was added to each sample and placed in the TissueLyser (Qiagen, Valencia, CA) for 5 minutes at 30 Hz at room temperature. RNA was extracted according to the manufacturer's instructions, and quantified using a spectrophotometer. For RNA extraction from pancreas, a small piece of the pancreas was placed in a tube containing 600ul of RLT buffer (with B-mercapto ethanol) (Qiagen, Valencia, CA), and homogenized using the PowerGen 125 (Fisher scientific, Pittsburgh, PA). Following homogenization, Qiagen RNeasy Mini Protocol for Isolation of Total RNA from Animal Tissues (Qiagen, Valencia, CA) was used. After extraction, samples with a 260/280 reading of less than 1.65 failed quality control and were reextracted using the TRIzol protocol. One hundred micrograms of RNA from samples that passed quality control were then cleaned up using the RNeasy mini protocol for RNA cleanup (Qiagen, Valencia, CA), including the optional Dnase step. Samples were then quantified using a spectrophotometer, and assessed using the RNA 6000 nano assay (Agilent

Technologies, Palo Alto, CA) on the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). Samples that did not pass quality control after being run on the Bioanalyzer were discarded.

## **DNA Extraction**

Parental DNAs were purchased from The Jackson Laboratory (Bar Harbor, ME). F<sub>1</sub> DNAs were extracted from either tail or liver biopsy using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma, St.Louis, MO) according to the manufacturers' recommendations.

## **Assay Design**

### **Primer design**

Primers were designed for both *Aldoa* and *Prodh*. An important feature of an amplicon is that it contains polymorphisms between the strains. If there are no variants between parental strains, it is impossible to determine which product is from which parent. In addition, primer binding sites had to be identical across all strains which would be tested with that assay. If the primers do not bind equally to all the strains tested, then the difference in the quantity of product could be simply due to differences in primer binding rather than differences in the quantity of template. Ideally, primers would also span an intron, so that when amplifying from RNA, any DNA contamination in the sample would not contribute to the signal. When primers could not meet all of these criteria including that they be intron spanning, non-intron spanning primers were used and the expected amount of DNA contamination was subtracted. The primers designed for *Aldoa* and *Prodh* follow:

*Aldoa*:            F: 5' ACTTGAAGCGCTGCCAGTATGTT 3'

R: 5' GTCACAGCACTTCGTCGCACAGTGCC 3'

*Prodh*: F: 5' TGATGGAGGTACTCCCTTACCTGT 3'

R: 5' AAACAGGCTTCCTAGTCTGTCCTC 3'

The *Aldoa* forward primer binding site is in exon 6 and the reverse primer is located in exon 7; together, the primer set spans intron 6-7 which is only 92bp long. The advantage of a small intron is that in addition to amplification from RNA, PCR from DNA is also possible. This ability to amplify from both allows one to test primer binding in DNA across species, which is not possible to do from RNA. The expected product sizes are 180bp or 272bp from an RNA or DNA template respectively, which can be easily distinguished by gel electrophoresis. In addition *Aldoa* is widely expressed, and had been sequenced in all strains to be used in this study by the Pardo Manuel de Villena lab. The availability of genomic sequence across all strain allows us to know for sure that there are no polymorphisms in the primer binding sites, and that the expected product sizes are not very different so they will be easily identified by gel electrophoresis. For these reasons, *Aldoa* was used as a test gene for all standardization steps.

The *Prodh* primers are not intron-spanning. The forward primer binding site is located in the last exon 14, and the reverse is located in the 3'UTR, and the total expected product size from DNA or RNA is 565bp. The reason for selecting *Prodh* is that it is the smallest candidate gene, and that there was enough variation between the parental strains in that amplicon so that many F<sub>1</sub>s could potentially be tested for DAE. The primer pairs were tested over DNA (B6 and F<sub>1</sub>) and RNA (F<sub>1</sub>), just to ensure that the primers work.

## PCR/RT-PCR

All PCR and RT-PCR reactions were carried out using Eppendorf cMaster RT<sub>plus</sub>PCR System (Westbury, NY), and <sup>32</sup>P labeled dCTP. The manufacturers' recommendations were followed with the exception of modified PCR conditions which follow:

*Aldoa*: Annealing temp:68°C, extension for 45 seconds, for 25 cycles.

*Prodh*: Annealing temp:68°C, extension for 30 seconds, for 30 cycles

**Quantification:** All of the imaging data was collected using the Typhoon 9400 Variable Mode Imager (Amersham Biosciences, Piscataway, NJ). Quantification was performed using the ImageQuant™ V 5.2 molecular dynamics software (Amersham Biosciences, Piscataway, NJ). Local average was subtracted for background correction.

**Single Strand Conformation Polymorphism (SSCP):** All Reactions were run on SSCP gels prepared using MDE gel solution (Cambrex, Baltimore, MD) according to the manufacturer's instructions.

## Standardization

The following standardization steps were required before assaying for DAE. These are essential experimental standardizations to ensure that any inference of DAE is accurate.

**Linear Range:** The first step in identifying DAE was to ensure that the PCR is in the linear phase – i.e., has not yet reached the plateau stage. To identify the linear range of the PCR, different amounts of DNA were used as a template, and separated by gel electrophoresis (SSCP) and quantified. The PCR conditions were systematically altered until plots of DNA mass vs. signal intensity were linear or nearly so. DNA quantities ranged from 2.5 ng to 200 ng. All reactions were done in duplicate.



**Equal amplification:** To ensure that a primer pair could amplify equally from all strains, DNA from all parental strains was amplified and quantified (as described above). If there were any strains that showed significantly lower or higher amplification, they were either excluded from the panel for that gene or a new primer set was designed.

**DNA Quantification:** Whenever non-intron spanning primers were used, one concern is the level of background DNA contamination. The level of DNA contamination following all RNA extraction steps was expected to be low according to the manufacturers' information – for this reason, it was not reasonable to use a spectrophotometer to quantify DNA contamination. To quantify the amount of DNA contamination, an intron-spanning primer pair for *Aldoa* was used in the RT-PCR reactions with varying quantities of starting material. Using a standard curve and the Typhoon imager, levels of DNA contamination in RNA samples were quantified, and the amount of DNA contamination across samples per 100 ng of RNA was determined.

## **Results**

RNA was extracted from three males and three females from each type of F<sub>1</sub>. The F<sub>1</sub> RNA samples that passed quality control, were available for further experiments, they are listed in Table 4.1. Prior to assessing the presence or absence of DAE for each gene, several standardization steps had to be performed. First an assay had to be designed according to the criteria previously mentioned. This involved designing primers and testing both with DNA and RNA templates. *Aldoa* was chosen as a test gene for many reasons that are outlined in the assay design section. Briefly, *Aldoa* was chosen as it is highly expressed in many tissues (including brain), and because of the availability of high-quality genomic sequence for the strains used in this experiment, intron-spanning primers that allow the differentiation

between PCR products made from a DNA template and those from an RNA template, and the size of RNA and DNA products could be easily resolved.

**Determination of the linear range:** The first step which all subsequent steps rely on is determination of the linear range of the PCR. The purpose of this standardization step is to identify the number of cycles beyond which the PCR has reached a plateau; if the PCR has reached a plateau, the ratio of the alleles is not constant, and no longer represents the true ratio of the alleles. To determine this range, different amounts B6 DNA were used to seed PCR reactions in combination with different numbers of PCR cycles. Products were subjected to SSCP gel electrophoresis, the gel was dried and exposed to a phosphor screen, and signal intensity was measured using the Typhoon imager (Figures 4.1 and 4.2). After correcting for background, signal intensities from each sample were plotted against the amount (ng) of template used, and the conditions that produced the regression line with the highest coefficient of correlation were considered the linear range. This was done for *Aldoa* and *Prodh* (Figures 4.3 and 4.4). The linear range for *Aldoa* and *Prodh* were determined to be 25 cycles and 0-25 ng template and 30 cycles and 0-30 ng template, respectively.

**Equal Amplification:** One of the concerns mentioned above was that of primer design. Since several different strains were used with a given set of primers, it was critical that the primers amplify all strains equally. If primers were bound more efficiently to one parental strain RNA than another, it would appear that there is DAE, when in reality this is an artifact of difference in primer binding. To test for this, DNA from all parental strains were amplified using a primer set, and the products were run on an SSCP gel and quantified. In order for an F<sub>1</sub> to be tested for DAE, both the parental strains had to amplify equally. The equity of amplification was assessed for both *Aldoa* and *Prodh*.

**DNA contamination:** One anticipated problem was the presence of DNA contamination in the RNA samples. The goal of this step was to determine the remaining amount of DNA contamination following extensive RNA cleanup steps for brain samples. This was done only using the *Aldoa* primers, because the amount of DNA contamination is a sample-specific feature and would be identical for all primers tested on a sample.

To accomplish this goal, the information from the linear range experiment was utilized for RT-PCR conditions. A standard curve was generated using B6 DNA in the range previously determined, and various amounts (ng) of RNA were amplified. The signal intensity from the Typhoon Imager for each band produced from DNA contamination was fit to the line generated by the standard, and the percent contamination (ng DNA/ng RNA) was plotted against the amount (ng) of RNA used (Figure 4.5). However, because the DNA contamination was very low for the smaller amounts of RNA, the readings were thought to be unreliable. For this reason the amount of contamination per 100 ng of RNA was determined based on the mode rather than the mean. The resulting contamination equaled 4 ng DNA per 100 ng RNA. Therefore, subsequent experiments using RNA extracted from brain were assumed to have the same level of DNA contamination, and in order for a gene to be tested for DAE the level of expression will have to be much higher than the equivalent of 4 ng of DNA.

**Differential Allelic Expression:** Following extensive standardization and optimization, DAE was assessed. Since *Aldoa*—a gene used for standardization—was of no interest to our study of genes relevant to schizophrenia, it was not tested for DAE. The only other gene that had been assessed for linear range and equal amplification aside from *Aldoa* in the brain was *Prodh*. To begin testing for DAE, one has to determine how much RNA template is to be

used in order to be within the linear range. Several different amounts of RNA were tested in relation to the standard curve (Figure 4.6). However, for *Prodh*, the expression was lower than what is expected for background from DNA contamination.

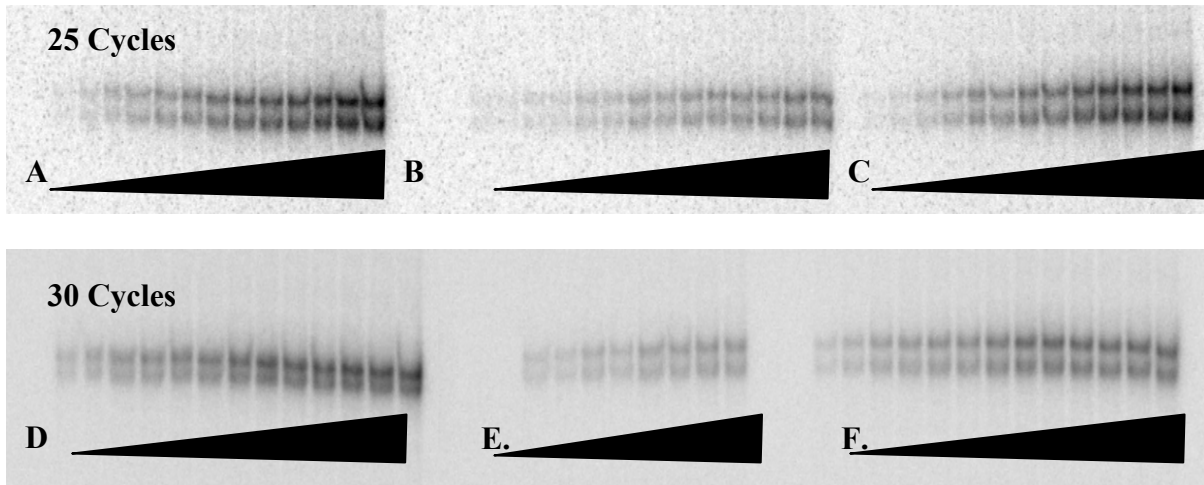
### **Conclusion:**

In this study several goals have been completed. The first aim was to create a set of F<sub>1</sub> RNA samples to be used to study DAE. For all F<sub>1</sub>s generated to date RNA samples from various tissues are available. These RNAs will serve as a tool for many future studies in the Pardo Manuel de Villena, Mohlke, and Sullivan labs. The second aim was to identify and standardize the steps necessary prior to testing for DAE. The steps required have been identified. They include the determination of the linear range of a PCR, level of DNA contamination, and equal amplification. For both the *Aldoa* and *Prodh* assays the linear range was determined. Expected DNA contamination across brain RNA samples was determined using the *Aldoa* assay. In addition both the *Aldoa* and *Prodh* assays were found to amplify equally from all parental strains. The final aim of this study was to assess the presence or absence of DAE of candidate genes. *PRODH* was the only gene tested for DAE, however its expression in brain was so low that differential expression could not be determined.

This study has laid the ground work for future experiments to assess DAE further. Necessary standardization steps including determination of the linear range, DNA contamination and equal amplification have been developed. While no genes were successfully tested for DAE, significant amount of knowledge on identifying DAE has been gained, and these samples and tools are available for on-going and future studies.

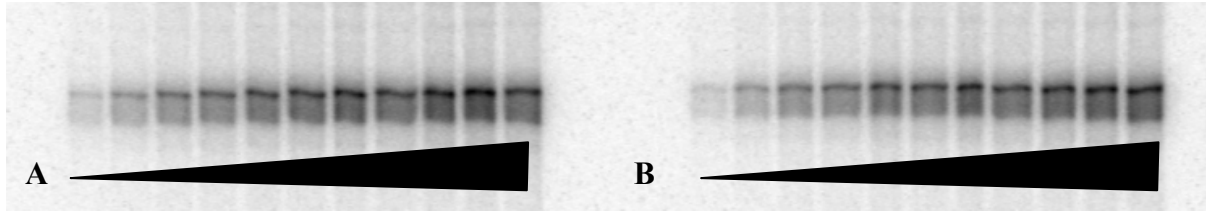
Selection of the panel of F<sub>1</sub>s and experimental design was completed by Fernando Pardo Manuel de Villena. All of the F<sub>1</sub>s were generated by Timothy A. Bell in the Pardo Manuel de Villena lab. All RNA samples from pancreas were extracted by Li Qin in the Mohlke lab. I extracted brain and cerebellum RNAs, and completed all the standardization steps.

**Figure 4.1**



**Figure 4.1:** AldoA Linear Range: **A.** 25-cycle PCR seeded with 0-200 ng B6 DNA. **C.** Replicate PCR of A. **B.** The first eight samples from a and c reloaded in the following order: 0a, 0c, 2.5a, 2.5c, 5a, 5c, 10a, 10c, 15a, 15c, 20a, 20c, 25a, 25c, 50a, 50c. **D.** 30-cycle PCR seeded with 0-200ng B6 DNA. **F.** Replicate PCR of D. in the same order. **E.** The first five samples from D and F reloaded in the following order: 0d, 0f, 2.5d, 2.5f, 5d, 5f, 10d, 10f, 15d, 15f. The black triangles show increasing quantity of DNA used to seed the reactions.

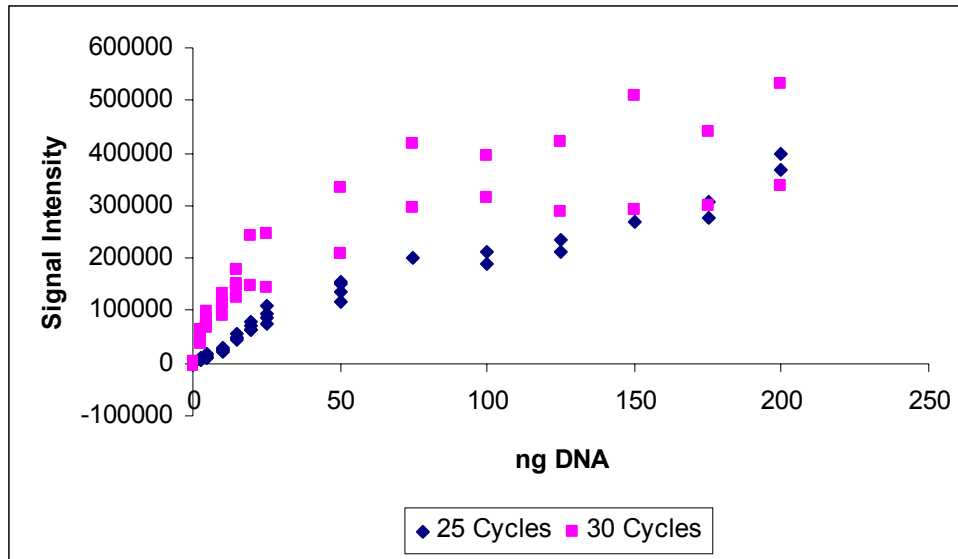
**Figure 4.2**



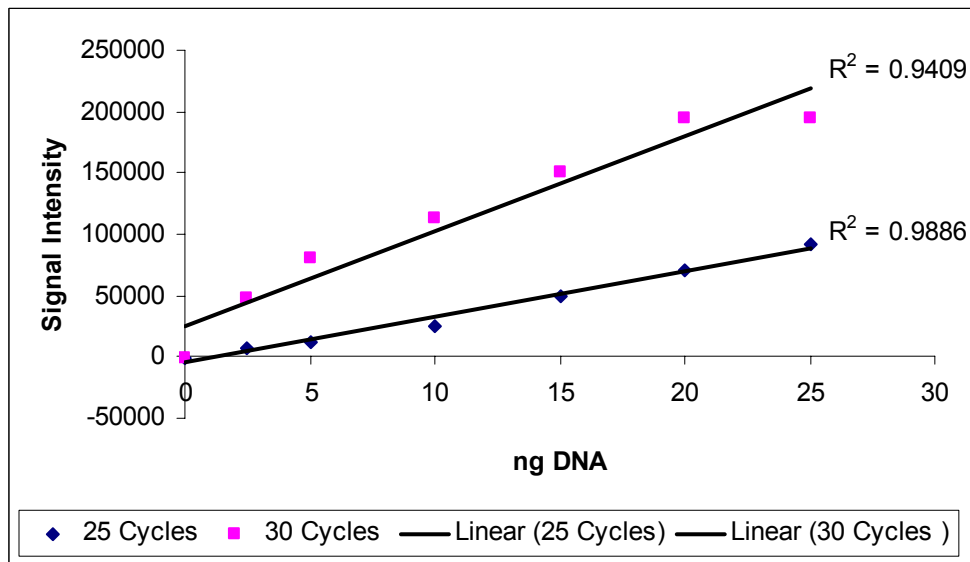
**Figure 4.2:** Prodh Linear Range: **A.** 30-cycle PCR seeded with 0-50ng B6 DNA. **B.** Replicate PCR of A. in the same order. The black triangles show increasing quantity of DNA used to seed the reactions.

**Figure 4.3**

**A.**



**B.**

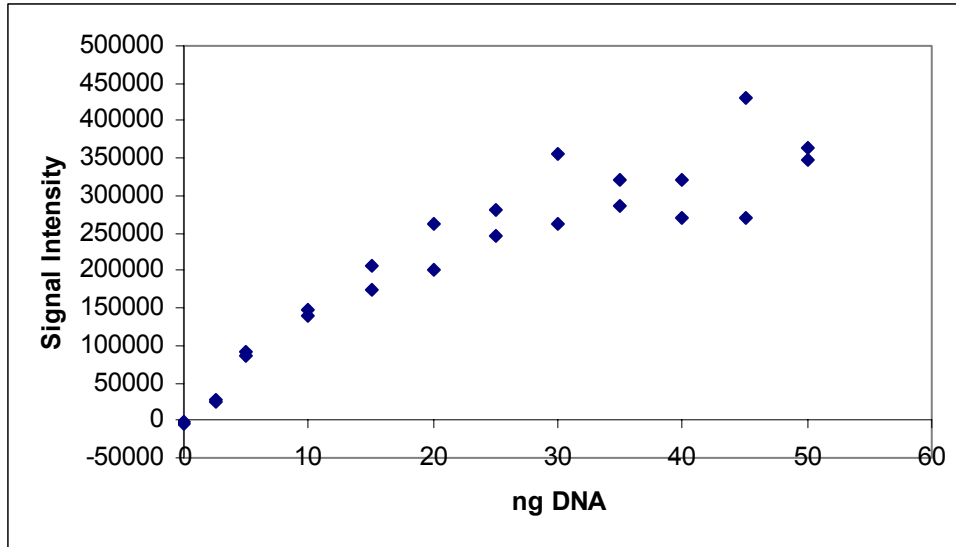


**Figure 4.3:** Graph of Aldoa Linear Range: **A.** Band intensities from the gels were quantified using a Typhoon imager, and plotted against the starting amount (ng) of DNA **B.** A plot of the average for each quantity of template, with the line of best fit. Following optimization experiments, 25 PCR cycles and 0 to 25 ng were determined to yield products in the linear phase of the PCR for this primer pair.

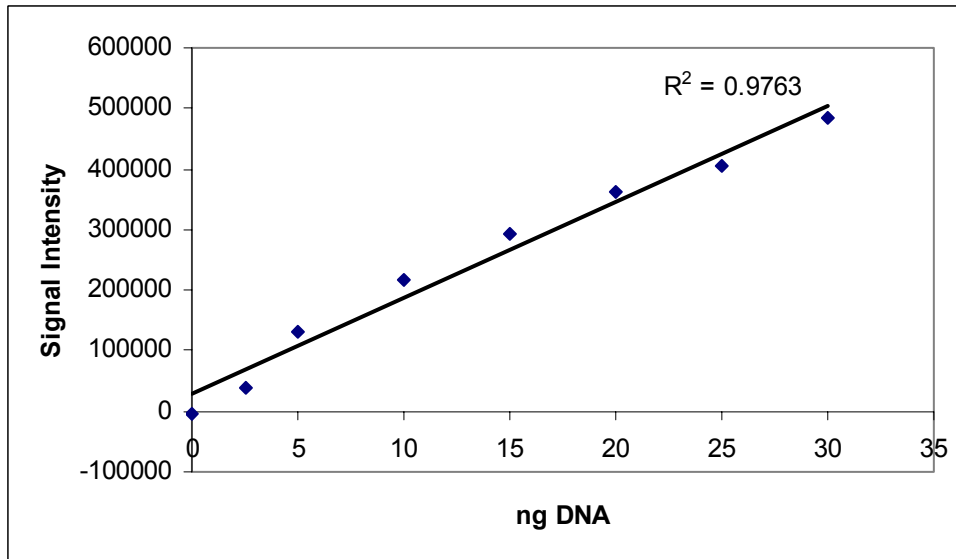


**Figure 4.4**

**A.**



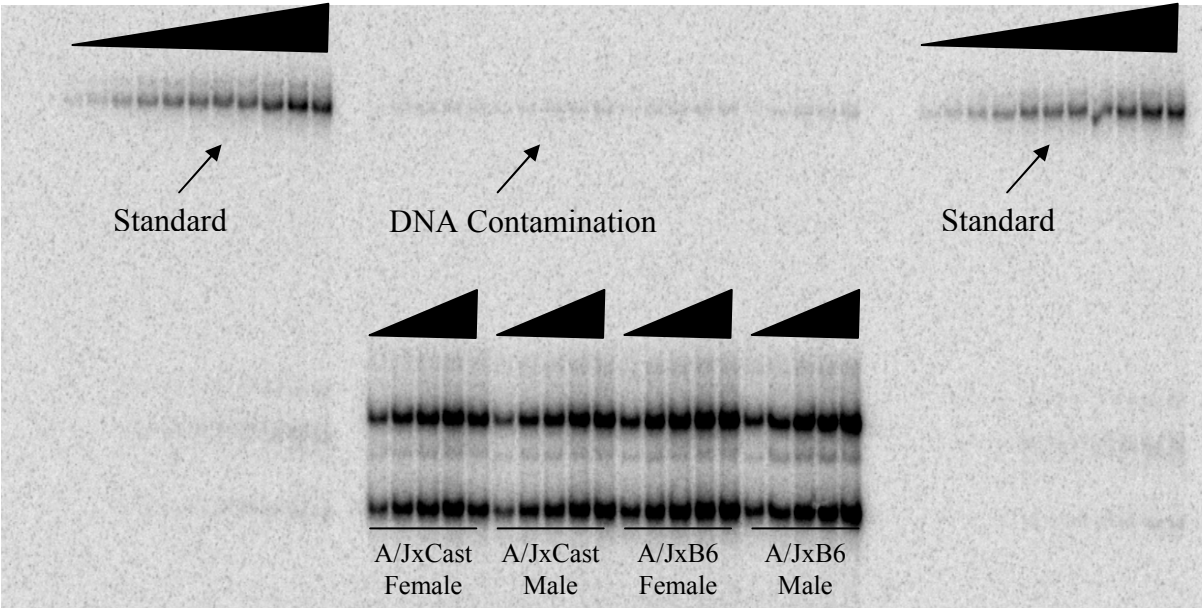
**B.**



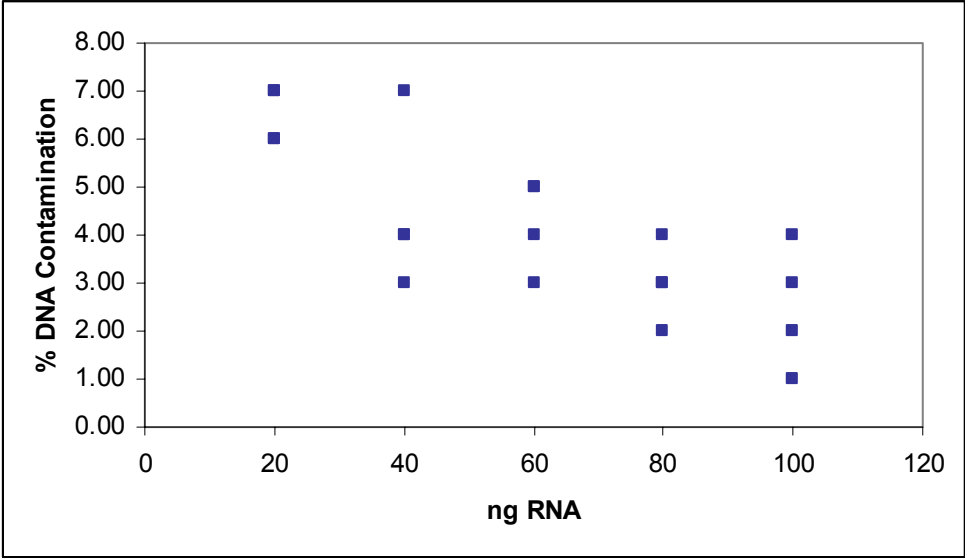
**Figure 4.4:** Graph of Prodh Linear Range: **A.** All points for 30 cycles. **B.** A plot of the average for each quantity of template, with the linear line of best fit. 30 cycles and 0 to 30 ng was determined to be the linear phase of the PCR for this primer pair.

Figure 4.5

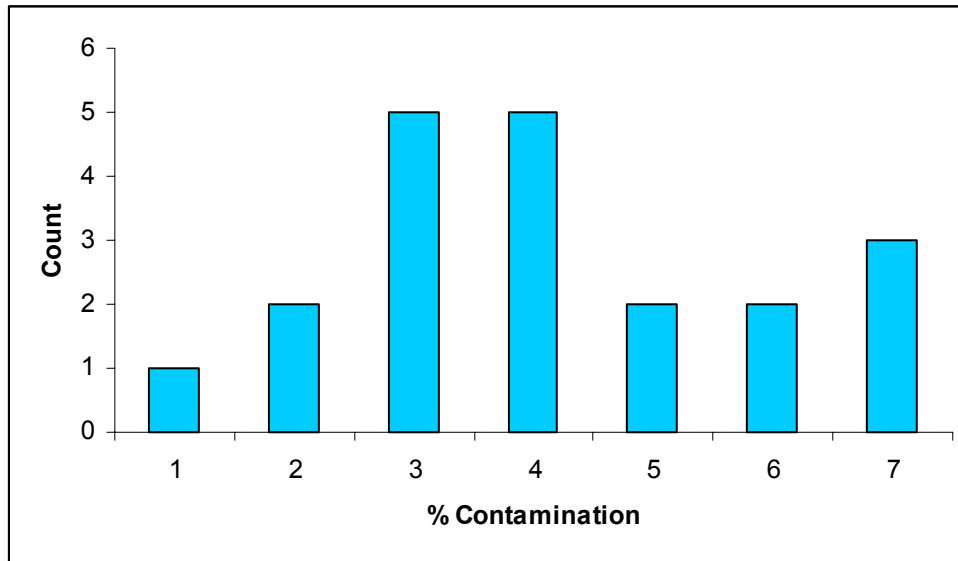
A.



B.

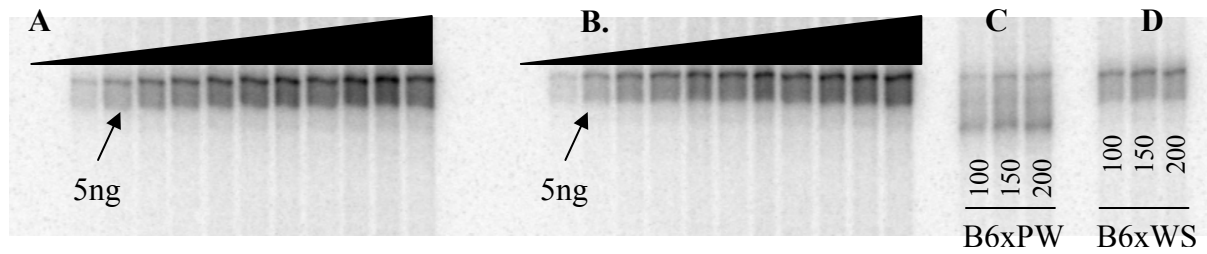


C.



**Figure 4.5:** DNA Contamination: **A.** *Aldoa* was used to determine the general amount of DNA contamination. Two standard curves were run using 0 – 50ng DNA. Quantity of DNA contamination was determined using 20, 40, 60, 80, or 100ng RNA from four different F<sub>1</sub>s. **B.** Percent DNA contamination vs. quantity of RNA used to seed the reaction. **C.** Percent DNA contamination plotted vs. the number of reactions that yield that quantity of contamination to determine the mode, which is about four percent or 4ng DNA per 100ng RNA. The black triangles show increasing quantity of DNA or RNA used to seed the reactions.

**Figure 4.6**



**Figure 4.6:** PROD<sup>H</sup> Expression: **A, B.** A standard curve for *Prodh* was loaded in duplicate from 0-50 ng DNA. **C, D.** One hundred to two hundred nanograms of F<sub>1</sub> RNA was used to seed the RT-PCR reactions. The total product is not more than what is expected to result from DNA contamination of about 4 ng DNA per 100 ng RNA, as can be seen in relation to the 5 ng band in the standard curve. This leads to the conclusion that DAE cannot be tested in *Prodh* using this primer pair.

**Table 4.1:**

F1	Sex	Brain	Cerebellum	Heart	Lung	Colon	Liver	Kidney	Ovary/Testis	Spleen	Pancreas
A/J x B6	Male	3/3	1/3	3	3	3	3	3	3	3	
	Female	3/3	0/3	3	3	3	3	3	3	3	
A/J x Cast	Male	1/1	0/1								1/1
	Female	3/3	0/3								3/3
B6 x A/J	Male	2/2	0/3	3	3	3	3	3	3/3	3	
	Female	3/3	1/3	3	3	3	3	3	3/3	3	
B6 x PWD	Male	3/3	0/3	3	3	3	3	3	3	3	
	Female										3/3
B6 x WSB	Male	2/2	0/3	3	3	3	3	3	3	3	3/3
	Female	3/3	0/3	3	3	3	3	3	3	3	3/3
B6 x PERA	Male	1/1	0/1	1	1	1	1	1	1	1	
	Female	3/3	0/3	3	3	3	3	3	3	3	
Cast x A/J	Male	2/2	0/3	3	3	3	3	3	3/3	3	
	Female	3/3	0/3	3	3	3	3	3	3/3	3	
FVB x B6	Male	3/3	0/3	3	3	3	3	3	3/3	3	3/3
	Female	3/3	0/3								3/3
FVB x PWD	Male	3/3	0/3								3/3
	Female	3/3	0/3								3/3
JF1 x B6	Male										
	Female	1/1	0/1	1	1	1	1	1	1	1	1/1
PERA x B6	Male	3/3	0/3	3	3	3	3	3	3/3	3	3/3
	Female	2/2	0/3	3	3	3	3	3	3/3	3	3/3
PERA x WSB	Male	2/2	0/2	2	2	2	2	2	2/2	2	2/2
	Female	1/1	0/3	3	3	3	3	3	3/3	3	1/1
PERC x B6	Male	1/1	0/1								1/1
	Female	3/3	0/3								3/3
PWD x A/J	Male	2/2	0/2	2	2	2	2	2	2/2	2	3/3
	Female	3/3	0/3	3	3	3	3	3	3/3	3	3/3
PWD x B6	Male	3/3	0/3	3	3	3	3	3	3/3	3	
	Female	3/3	0/3	3	3	3	3	3	3/3	3	
PWD x WSB	Male										2/2
	Female										
RBA x A/J	Male	2/2	0/2	2	2	2	2	2	2/2	2	1/2
	Female	1/1	0/1	1	1	1	1	1	1/1	1	1/1
WSB x B6	Male	3/3	0/3	3	3	3	3	3	3	3	
	Female	3/3	0/3	3	3	3	3	3	3	3	
WSB x PERC	Male										
	Female										2/2

**Table 4.1:** RNA samples available: All RNA samples which passed quality control are included in this table. The ratios listed are the number of samples either tested for DNA contamination or used for other experiments out of the total available for each cross. Whole numbers represent the total number of samples available, and are currently being updated.

## Chapter 5

### ***Conclusion***

The studies discussed in this thesis are aimed to investigate genetic variation in candidate genes for schizophrenia research. The goal of the study discussed in Chapter 2 “Genotyping Accuracy for Whole-Genome Amplification of DNA from Buccal Epithelial Cells” was to determine if the GenomiPhi DNA Amplification Kit accurately amplified buccal cell DNA for various markers. Several goals were accomplished. Upon evaluation of the amplification kit using a rolling circle/ $\Phi$ 29 DNA polymerase, it was determined to be sufficiently accurate for the markers tested. In future studies, markers would first need to be tested in both unamplified and amplified samples to determine accuracy for that region. While independent testing will be necessary for other markers, this work shows that the kit used performs better than other commercially available whole-genome amplification kits. In the future it could be used to amplify samples for which very little DNA is available.

In Chapter 3, the goal was to determine the degree of association between genetic variation in *COMT* and neurocognition or working memory. Chapter 3 of this work contained the largest case-only association study of *COMT* SNPs with neurocognition or working memory. The results of this study suggest that *COMT* genotype alone is not related to neurocognition. However the study was designed prior to the release of the most recent HapMap build. Better SNP coverage of *COMT* may yield different results. Lack of association also does not exclude association of *COMT* expression changes, methylation

pattern or other alterations with neurocognitive function. This clearly requires further investigation. SNPs in other candidate genes described in Chapter 1 (*AKT1*, *DISC1/DISC2*, *SLC6A3*, *DRD3*, *DTNBP1*, *NRG1*, *PRODH*, *RGS4*, and *SLC5A4*) will also need to be genotyped and tested for association with neurocognition. In addition, multiple gene interactions should be assessed in relation to neurocognitive phenotype since multiple genes could have small main effects individually, but a large combined effect. Appropriate controls should also be included in future studies.

In Chapter 4, an attempt was made to identify differential expression of candidate genes for schizophrenia. An approach to studying differential expression of candidate genes was detailed. The process by which DAE can be determined was elucidated, and the standardization steps were performed for two genes, *Aldoa* and *Prodh*. No genes were actually found to exhibit differential expression in mouse brain. The method outlined in Chapter 4, can prove be a very powerful tool in identification of differential expression and in the determination of the alleles responsible for the changes in expression. These experiments and the standardization protocols established here will facilitate DAE assessment for other candidate genes listed above, and in Chapter 1. In addition these samples are likely to prove valuable for many other studies.

The studies discussed in this thesis have laid the groundwork for future experiments studying the genetics of schizophrenia.

## Works Cited

- Addington, J., D. Addington, et al. (1991). "Cognitive functioning and positive and negative symptoms in schizophrenia." Schizophr Res **5**(2): 123-34.
- Alfimova, M. and L. Uvarova (2003). "Cognitive peculiarities in relatives of schizophrenic and schizoaffective patients: heritability and resting EEG-correlates." Int J Psychophysiol **49**(3): 201-16.
- American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, American Psychiatric Association.
- Anokhin, A. P., A. C. Heath, et al. (2003). "Genetic influences on frontal brain function: WCST performance in twins." Neuroreport **14**(15): 1975-8.
- Association, A. P. (1994). Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, American Psychiatric Association.
- Barrett, J. C., B. Fry, et al. (2004). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics.
- Bellgrove, M. A., K. Domschke, et al. (2005). "The methionine allele of the COMT polymorphism impairs prefrontal cognition in children and adolescents with ADHD." Exp Brain Res **163**(3): 352-60.
- Bilder, R. M., R. S. Goldman, et al. (2000). "Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates." Am J Psychiatry **157**(4): 549-59.
- Bouchard, T. J., Jr. (1998). "Genetic and environmental influences on adult intelligence and special mental abilities." Hum Biol **70**(2): 257-79.
- Bray, N. J., P. R. Buckland, et al. (2003). "A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain." Am J Hum Genet **73**(1): 152-61.



- Camp, N. J., S. L. Neuhausen, et al. (2001). "Genomewide multipoint linkage analysis of seven extended Palauan pedigrees with schizophrenia, by a Markov-chain Monte Carlo method." Am J Hum Genet **69**(6): 1278-89.
- Cannon, T. D., M. O. Huttunen, et al. (2000). "The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia." Am J Hum Genet **67**(2): 369-82.
- Cannon, T. D., J. Kaprio, et al. (1998). "The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study." Arch Gen Psychiatry **55**(1): 67-74.
- Chen, J., B. K. Lipska, et al. (2004). "Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain." Am J Hum Genet **75**(5): 807-21.
- Chen, W. J., S. K. Liu, et al. (1998). "Sustained attention deficit and schizotypal personality features in nonpsychotic relatives of schizophrenic patients." Am J Psychiatry **155**(9): 1214-20.
- Chen, X. and P. F. Sullivan (2003). "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost, and throughput." Pharmacogenomics Journal **3**: 77-96.
- Colhoun, H. M., P. M. McKeigue, et al. (2003). "Problems of reporting genetic associations with complex outcomes." Lancet **361**(9360): 865-72.
- Cornblatt, B. A., N. J. Risch, et al. (1988). "The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families." Psychiatry Res **26**(2): 223-38.
- Cowles, C. R., J. N. Hirschhorn, et al. (2002). "Detection of regulatory variation in mouse genes." Nat Genet **32**(3): 432-7.
- Dean, F. B., S. Hosono, et al. (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A **99**(8): 5261-6.
- Devlin, B. and N. Risch (1995). "A comparison of linkage disequilibrium measures for fine-scale mapping." Genomics **29**: 311-22.

- Devlin, B. and K. Roeder (1999). "Genomic control for association studies." Biometrics **55**: 997-1004.
- Efron, B. and R. Tibshirani (1993). An Introduction to the Bootstrap. London, Chapman and Hall.
- Egan, M. F., T. E. Goldberg, et al. (2001). "Relative risk for cognitive impairments in siblings of patients with schizophrenia." Biol Psychiatry **50**(2): 98-107.
- Egan, M. F., T. E. Goldberg, et al. (2001). "Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia." Proc Natl Acad Sci U S A **98**(12): 6917-22.
- Elvevag, B. and T. E. Goldberg (2000). "Cognitive impairment in schizophrenia is the core of the disorder." Crit Rev Neurobiol **14**(1): 1-21.
- Faraone, S. V., L. J. Seidman, et al. (1999). "Neuropsychological functioning among the nonpsychotic relatives of schizophrenic patients: a 4-year follow-up study." J Abnorm Psychol **108**(1): 176-81.
- Farmer, A. E., P. McGuffin, et al. (1987). "Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition." Arch Gen Psychiatry **44**(7): 634-41.
- First, M., R. Spitzer, et al. (1994). Structured Clinical Interview for DSM-IV Axis I Disorders--Administration Booklet. Washington, DC, American Psychiatric Press, Inc.
- Francks, C., S. E. Fisher, et al. (2003). "Familial and genetic effects on motor coordination, laterality, and reading-related cognition." Am J Psychiatry **160**(11): 1970-7.
- Franke, P., W. Maier, et al. (1992). "Wisconsin Card Sorting Test: an indicator of vulnerability to schizophrenia?" Schizophr Res **6**(3): 243-9.
- Freedman, M. L., D. Reich, et al. (2004). "Assessing the impact of population stratification on genetic association studies." Nat Genet **36**(4): 388-93.
- Good, P. I. (2004). Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses. New York, Springer-Verlag.

- Gorroochurn, P., S. E. Hodge, et al. (2004). "Effect of population stratification on case-control association studies. II. False-positive rates and their limiting behavior as number of subpopulations increases." Hum Hered **58**(1): 40-8.
- Gottesman, II and T. D. Gould (2003). "The endophenotype concept in psychiatry: etymology and strategic intentions." Am J Psychiatry **160**(4): 636-45.
- Green, M. F. (1996). "What are the functional consequences of neurocognitive deficits in schizophrenia?" Am J Psychiatry **153**(3): 321-30.
- Harrison, P. J. and D. R. Weinberger (2005). "Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence." Mol Psychiatry **10**(1): 40-68; image 5.
- Heiman, G. A., S. E. Hodge, et al. (2004). "Effect of population stratification on case-control association studies. I. Elevation in false positive rates and comparison to confounding risk ratios (a simulation study)." Hum Hered **58**(1): 30-9.
- Heinrichs, R. W. and K. K. Zakzanis (1998). "Neurocognitive deficit in schizophrenia: a quantitative review of the evidence." Neuropsychology **12**(3): 426-45.
- Hoff, A. L., H. Riordan, et al. (1992). "Neuropsychological functioning of first-episode schizophreniform patients." Am J Psychiatry **149**(7): 898-903.
- Horowitz, A., S. Shifman, et al. (2005). "A survey of the 22q11 microdeletion in a large cohort of schizophrenia patients." Schizophr Res **73**(2-3): 263-7.
- Hosono, S., A. F. Faruqi, et al. (2003). "Unbiased whole-genome amplification directly from clinical samples." Genome Res **13**(5): 954-64.
- Jablensky, A., N. Sartorius, et al. (1992). "Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study." Psychol Med Monogr Suppl **20**: 1-97.
- Kang, A. M., M. A. Palmatier, et al. (1999). "Global variation of a 40-bp VNTR in the 3'-untranslated region of the dopamine transporter gene (SLC6A3)." Biol Psychiatry **46**(2): 151-60.

- Keefe, R. S., R. M. Bilder, et al. (Submitted). "Baseline neurocognitive deficits in the CATIE schizophrenia trial."
- Keefe, R. S., R. C. Mohs, et al. (2003). "Neurocognitive assessment in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project schizophrenia trial: development, methodology, and rationale." Schizophr Bull **29**(1): 45-55.
- Keefe, R. S., J. M. Silverman, et al. (1994). "Performance of nonpsychotic relatives of schizophrenic patients on cognitive tests." Psychiatry Res **53**(1): 1-12.
- Kety, S. S. (1988). "Schizophrenic illness in the families of schizophrenic adoptees: findings from the Danish national sample." Schizophr Bull **14**(2): 217-22.
- Kety, S. S., P. H. Wender, et al. (1994). "Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark." Arch Gen Psychiatry **51**(6): 442-55.
- Knight, J. C. (2004). "Allele-specific gene expression uncovered." Trends Genet **20**(3): 113-6.
- Kremen, W. S., S. V. Faraone, et al. (1998). "Neuropsychological risk indicators for schizophrenia: a preliminary study of female relatives of schizophrenic and bipolar probands." Psychiatry Res **79**(3): 227-40.
- Laurent, A., O. Moreaud, et al. (1999). "Neuropsychological functioning among non-psychotic siblings and parents of schizophrenic patients." Psychiatry Res **87**(2-3): 147-57.
- Lazzeroni, L. C. and K. Lange (1998). "A conditional inference framework for extending the transmission/disequilibrium test." Hum Hered **48**(2): 67-81.
- Lesch, K. P., D. Bengel, et al. (1996). "Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region." Science **274**(5292): 1527-31.
- Leviel, K., M. Olarte, et al. (2004). "Genotyping accuracy for whole-genome amplification of DNA from buccal epithelial cells." Twin Research **7**: 482-4.

- Lieberman, J. A., T. S. Stroup, et al. (2005). "Effectiveness of antipsychotic drugs in patients with chronic schizophrenia." N Engl J Med **353**: 1209-23.
- Little, J., L. Bradley, et al. (2002). "Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations." Am J Epidemiol **156**(4): 300-10.
- Livak, K. J. (1999). "Allelic discrimination using fluorogenic probes and the 5' nuclease assay." Genet Anal **14**(5-6): 143-9.
- Luciano, M., M. J. Wright, et al. (2004). "A genetic investigation of the covariation among inspection time, choice reaction time, and IQ subtest scores." Behav Genet **34**(1): 41-50.
- Manschreck, T. C., B. A. Maher, et al. (1985). "Deficient motor synchrony in schizophrenic disorders: clinical correlates." Biol Psychiatry **20**(9): 990-1002.
- Michie, P. T., A. Kent, et al. (2000). "Phenotypic markers as risk factors in schizophrenia: neurocognitive functions." Aust N Z J Psychiatry **34 Suppl**: S74-85.
- Mills, S., K. Langley, et al. (2004). "No evidence of association between Catechol-O-Methyltransferase (COMT) Val158Met genotype and performance on neuropsychological tasks in children with ADHD: a case-control study." BMC Psychiatry **4**: 15.
- Moldin, S. O. and Gottesman, II (1997). "At issue: genes, experience, and chance in schizophrenia--positioning for the 21st century." Schizophr Bull **23**(4): 547-61.
- Myles-Worsley, M. and H. Coon (1997). "Genetic and developmental factors in spontaneous selective attention: a study of normal twins." Psychiatry Res **71**(3): 163-74.
- Nelson, J. R., Y. C. Cai, et al. (2002). "TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing." Biotechniques Suppl: 44-7.
- Nuechterlein, K. H., D. M. Barch, et al. (2004). "Identification of separable cognitive factors in schizophrenia." Schizophr Res **72**(1): 29-39.
- Oliphant, A., D. L. Barker, et al. (2002). "BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping." Biotechniques Suppl: 56-8, 60-1.

- Onstad, S., I. Skre, et al. (1991). "Twin concordance for DSM-III-R schizophrenia." Acta Psychiatr Scand **83**(5): 395-401.
- Palmatier, M. A., A. J. Pakstis, et al. (2004). "COMT haplotypes suggest P2 promoter region relevance for schizophrenia." Mol Psychiatry **9**(9): 859-70.
- Park, S., P. S. Holzman, et al. (1995). "Spatial working memory deficits in the relatives of schizophrenic patients." Arch Gen Psychiatry **52**(10): 821-8.
- Pastinen, T., B. Ge, et al. (2005). "Mapping common regulatory variants to human haplotypes." Hum Mol Genet **14**(24): 3963-71.
- R Development Core Team (2005). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.
- Risch, N. and J. Teng (1998). "The relative power of family-based and case-control designs for linkage disequilibrium studies of human complex traits: I. DNA pooling." Genome Research **8**: 1273-1288.
- Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**: 847-56.
- Roberts, R., P. Joyce, et al. (2000). "Rapid and comprehensive determination of cytochrome P450 CYP2D6 poor metabolizer genotypes by multiplex polymerase chain reaction." Hum Mutat **16**(1): 77-85.
- Rosa, A., V. Peralta, et al. (2004). "New evidence of association between COMT gene and prefrontal neurocognitive function in healthy individuals from sibling pairs discordant for psychosis." Am J Psychiatry **161**(6): 1110-2.
- Roxborough, H., W. J. Muir, et al. (1993). "Neuropsychological and P300 abnormalities in schizophrenics and their relatives." Psychol Med **23**(2): 305-14.
- Saha, S., D. Chant, et al. (2005). "A systematic review of the prevalence of schizophrenia." PLoS Med **2**(5): e141.
- SAS Institute Inc. (2004). SAS/STAT® Software: Version 9. Cary, NC, SAS Institute, Inc.

- Saykin, A. J., R. C. Gur, et al. (1991). "Neuropsychological function in schizophrenia. Selective impairment in memory and learning." Arch Gen Psychiatry **48**(7): 618-24.
- Saykin, A. J., D. L. Shtasel, et al. (1994). "Neuropsychological deficits in neuroleptic naive patients with first-episode schizophrenia." Arch Gen Psychiatry **51**(2): 124-31.
- Schaid, D. J., C. M. Rowland, et al. (2002). "Score tests for association between traits and haplotypes when linkage phase is ambiguous." Am J Hum Genet **70**(2): 425-34.
- Shields, A. E., M. Fortun, et al. (2005). "The use of race variables in genetic studies of complex traits and the goal of reducing health disparities." American Psychologist **60**: 77-103.
- Shifman, S., M. Bronstein, et al. (2002). "A highly significant association between a COMT haplotype and schizophrenia." American Journal of Human Genetics **71**: 1296-302.
- Shifman, S., M. Bronstein, et al. (2002). "A highly significant association between a COMT haplotype and schizophrenia." Am J Hum Genet **71**(6): 1296-302.
- Spitzer, M. (1993). "The psychopathology, neuropsychology, and neurobiology of associative and working memory in schizophrenia." Eur Arch Psychiatry Clin Neurosci **243**(2): 57-70.
- Stefansson, H., E. Sigurdsson, et al. (2002). "Neuregulin 1 and susceptibility to schizophrenia." Am J Hum Genet **71**(4): 877-92.
- Strauss, M. E. (1993). "Relations of symptoms to cognitive deficits in schizophrenia." Schizophr Bull **19**(2): 215-31.
- Stroup, T. S., J. P. McEvoy, et al. (2003). "The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development." Schizophr Bull **29**(1): 15-31.
- Stumvoll, M., O. Tschritter, et al. (2002). "Association of the T-G polymorphism in adiponectin (exon 2) with obesity and insulin sensitivity: interaction with family history of type 2 diabetes." Diabetes **51**(1): 37-41.
- Sullivan, P. F. (2005). "The genetics of schizophrenia." PLoS Med **2**(7): e212.

- Sullivan, P. F., L. J. Eaves, et al. (2001). "Genetic case-control association studies in neuropsychiatry." Arch Gen Psychiatry **58**: 1015-24.
- Sullivan, P. F., K. S. Kendler, et al. (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Arch Gen Psychiatry **60**(12): 1187-92.
- Swan, G. E. and D. Carmelli (2002). "Evidence for genetic mediation of executive control: a study of aging male twins." J Gerontol B Psychol Sci Soc Sci **57**(2): P133-43.
- Swan, G. E., D. Carmelli, et al. (1990). "Heritability of cognitive performance in aging twins. The National Heart, Lung, and Blood Institute Twin Study." Arch Neurol **47**(3): 259-62.
- Taerk, E., N. Grizenko, et al. (2004). "Catechol-O-methyltransferase (COMT) Val108/158 Met polymorphism does not modulate executive function in children with ADHD." BMC Med Genet **5**: 30.
- Tranah, G. J., P. J. Lescault, et al. (2003). "Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies." Biotechnol Lett **25**(13): 1031-6.
- Tunbridge, E., P. W. Burnet, et al. (2004). "Catechol-o-methyltransferase (COMT) and proline dehydrogenase (PRODH) mRNAs in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and major depression." Synapse **51**(2): 112-8.
- Turic, D., H. Williams, et al. (2005). "A family based study of catechol-O-methyltransferase (COMT) and attention deficit hyperactivity disorder (ADHD)." Am J Med Genet B Neuropsychiatr Genet **133**(1): 64-7.
- Tuulio-Henriksson, A., J. Haukka, et al. (2002). "Heritability and number of quantitative trait loci of neurocognitive functions in families with schizophrenia." Am J Med Genet **114**(5): 483-90.
- Wacholder, S., N. Rothman, et al. (2000). "Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias." J Natl Cancer Inst **92**: 1151-8.



- Weickert, C. S., R. E. Straub, et al. (2004). "Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain." Arch Gen Psychiatry **61**(6): 544-55.
- Wu, E. Q., H. G. Birnbaum, et al. (2005). "The economic burden of schizophrenia in the United States in 2002." J Clin Psychiatry **66**(9): 1122-9.
- Yan, H., Z. Dobbie, et al. (2002). "Small changes in expression affect predisposition to tumorigenesis." Nat Genet **30**(1): 25-6.
- Yang, W. S., P. L. Tsou, et al. (2003). "Allele-specific differential expression of a common adiponectin gene polymorphism related to obesity." J Mol Med **81**(7): 428-34.